THE UNIVERSITY *of* EDINBURGH

The molecular function of the RNA-binding protein NANOS2 in mouse spermatogonial stem cells

Azzurra Codino

Thesis presented for the degree of Doctor of Philosophy

Institute for Regeneration and Repair
Centre for Regenerative Medicine
The University of Edinburgh
2019

Declaration

I declare that this thesis has been composed by myself. The work contained in this thesis is my own, except where stated otherwise by reference or acknowledgment. This work has not been submitted for any other previous degree or professional qualification.

Azzurra Codino

Edinburgh, 2nd September 2019

To my grandmother.

A mia nonna.

“Consider ye the seed from which ye sprang;
Ye were not made to live like unto brutes,
But for pursuit of virtue and of knowledge”.

Divine Comedy, Hell, Canto 26th, Dante Alighieri

“Considerate la vostra semenza:
fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza”.

La Divina Commedia, Inferno, Canto 26mo, vv. 118-120, Dante Alighieri

Acknowledgments

I would like to express my gratitude to my supervisor, Prof. Dónal O'Carroll, for mentoring me during these four years and for having supported my work continuously. I have learned from Dónal a rigorous science, and a meticulous and analytical approach for the interpretation of experimental results. I am grateful to Dónal also for giving me the opportunity of learning many advanced and interesting techniques, from both his laboratory and collaborators.

I am also grateful to the members of my Thesis Committee meeting, Prof. Ian Chambers, Doctor Simon Tomlinson, and Prof. David Tollervey, for providing critical feedback on my project, during the past four years.

I would like to acknowledge also Prof. David Tollervey for allowing me learning CRAC from his lab and for the scientific discussions about my project. I would like to thank Dr. Tomasz Turowski, who has taught me all the secrets of CRAC, and the other members of the Tollevery lab for the beneficial discussions about my project.

I also would like to thank the laboratory of Prof. Juri Rappsilber for performing mass spectrometry.

A special thank also to all the previous and present members of the O'Carroll lab. I am grateful to Dr. Lina Vasiliauskaite for teaching me SSCs derivation and how to work with mice. I also would like to thank Dr. Monica Di Giacomo and Dr. Claudia Carrieri for teaching me the biology of spermatogonia, and Dr. Ivayla Ivanova for helping me with some experiments. I would like to thank Dr. Christian Much for the preparation of samples for mass spectrometry, and Dr. Ansgar Zoch for helping me with the IP-MS analysis. I also would like to thank Dr. Marcos Morgan and Dr. Francesco Severi for their feedback on my project and the scientific discussions, and Dr. Louie van de Lagemaat for his help with

the bioinformatic analysis. Moreover, I would like to thank Dr. Yuka Kabayama and Azzurra De Pace who have always helped me in the lab and supported me during my doctoral studies.

I would like to thank also all the people working in the facilities in my institute for their technical support and, in particular, Dr. Fiona Rossi for her special kindness, while helping me with FACS late in the night, during my time-course experiments.

I am grateful to Dr. Raphaël Pantier for his valuable explanations concerning the bioinformatic analysis, and for his constant support and patience during my doctoral studies.

I am also grateful to all those people I have met during my pre-doctoral and doctoral studies, who have helped me to discover my ambitions and encouraged me to pursue them.

Lay summary

Life is perpetuated by the reproduction of fertile organisms. Notably, adult male mammals are fertile throughout their life and can generate up to hundreds of million sperm daily. But how can they ensure such continuous sperm production? Which are the biological processes underlying their fertility? In mammals, sperm is produced after a long and highly organised developmental process called spermatogenesis, which starts from particular cells: spermatogonial stem cells (SSCs). SSCs constitute the endless source which fuels spermatogenesis, as they have the unique ability to both self-renew and differentiate. When they self-renew, they give rise to an identical SSC; whereas, upon differentiation, they generate more specialised cells, which ultimately undertake spermatogenesis and produce mature sperm. SSCs can be identified in adult testes by morphological features, molecular markers, as well as by functional assays in mouse models. However, the precise mechanisms underlying their self-renewal ability are poorly understood.

During my Ph.D., I used the mouse as a model organism to investigate the molecular processes which occur inside SSCs. Notably, I focused on NANOS2, a protein which is essential for SSCs self-renewal, and male mice fertility. By using a genetic modified mouse which expresses a variant of NANOS2 (TAG-NANOS2), I was able to 'fish' NANOS2 protein inside SSCs and to characterise which are its interacting partners, at the molecular level. Notably, I discovered that NANOS2 strongly associates with ~1400 mRNAs. These are recognised by NANOS2 because they have a particular motif, which distinguishes them from the rest of the transcriptome. Additionally, these mRNAs produce proteins involved in metabolic processes. I also discovered that these mRNAs are rapidly degraded in SSCs. Since NANOS2 cooperates with other proteins, called CNOTs, their interplay might be the cause of mRNAs decay. Overall, these findings constitute a great advance for the comprehension of NANOS2 molecular function, in mouse SSCs.

Importantly, also human NANOS1 was implicated in male infertility. Thus, as in mouse, it is possible that also human NANOS proteins repress mRNAs in healthy adult testes. In conclusion, my data could also pave the way for understanding the molecular basis of male infertility in human.

Abstract

Spermatogonial stem cells (SSCs) are adult unipotent stem cells which are able to both self-renew and differentiate to support spermatogenesis throughout adult life. NANOS2 is a conserved RNA-binding protein, and its genetic deletion in adult mice leads to male infertility due to loss of SSCs. However, the precise mechanism underlying NANOS2 function during SSCs maintenance remains largely unexplored at the transcriptomic level. To accurately identify NANOS2 RNA-occupancy, we generated a *Nanos2*^{TAG} mouse model, I derived and expanded SSCs *in vitro* and performed CRAC (UV-Cross-linking and Analysis of cDNA). CRAC revealed that NANOS2 strongly and specifically interacts with RNA, despite its low abundance in cells. It also showed that NANOS2 mainly binds mRNAs within their 3' UTRs, which are enriched in the novel 'AUNAANU' motif. These mRNAs are involved in cellular metabolism, and they also include positive regulators of the mTORC pathway, which is known to control SSCs self-renewal. Additionally, we analysed the expression profile of these mRNAs throughout adult spermatogonia development. This revealed that one-third of NANOS2 targets are upregulated during the transition from self-renewing SSCs to transit-amplifying spermatogonia.

To further explore NANOS2 function, we performed immunoprecipitation followed by mass spectrometry and identified components of the CCR4-NOT complex amongst the strongest interactors. Since the CCR4-NOT promotes RNA decay, I investigated the mRNA stability of NANOS2 targets by using SLAM-seq, a novel technique that enables the determination of mRNA half-life transcriptome-wide. This analysis suggested that NANOS2 mRNA targets have a shorter half-life, compared to the average mRNA stability in SSCs.

In summary, I demonstrated that NANOS2 shortens the half-life of hundreds of mRNAs involved in metabolism, through the recruitment of the CCR4-NOT RNA deadenylation complex, in SSC lines. Therefore, we speculate that NANOS2 promotes SSCs self-renewal by repressing active metabolism and

cellular growth. This would preserve cellular quiescence and ensures the presence of a long-term SSCs pool in adult mice. Overall, these findings provide new insight into NANOS2 mechanism of action in mouse SSC.

Table of Contents

List of figures

Abbreviations

5', 3' UTR. 5', 3' Untranslated region

bp. Base pair

CDS. Coding sequence

CLIP. UV-cross-linking and immunoprecipitation

Co-IP. Co-immunoprecipitation

CRAC. UV-cross-linking and analysis of cDNA

CRISPR. Clustered regularly interspaced short palindromic repeats

ESCs. Embryonic stem cells

EGF. Epidermal growth factor

FACS. Fluorescent-activated cell sorting

FGF2. Fibroblast growth factor

GDNF. Glial cell line-derived neurotrophic factor

Kb. Kilobase

kDa. Kilodalton

KO. Knockout

LIF. Leukemia inhibitory factor

MEFs. Mouse embryonic fibroblasts

MEME. Multiple EM for Motif Elicitation

mRNPs. messenger ribonucleoproteins particles

PBs. Processing bodies

PBS. Phosphate Buffered Saline

PCR. Polymerase chain reaction

PGCs. Primordial germ cells

RBP. RNA binding protein

Rpm. Revolutions per minute

R.T. Room temperature

s^4U. 4-thiouridine

SDS-PAGE. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

SSCs. Spermatogonial stem cells

WT. Wild-type.

Chapter 1. Introduction

1.1 Mouse spermatogenesis

The precursors of male germ cells originate during early embryonic development from Primordial Germ Cells (PGCs), which arise from the epiblast, as a result of inductive signals (Hayashi et al., 2007). PGCs induction occurs at E5.5-6.25 and requires Bmp4 signalling from the extraembryonic ectoderm (Lawson et al., 1999; Ohinata et al., 2009) (Figure 1.1 A). PGCs specification also necessitates the expression of different transcription factors, including Blimp1 (Ohinata et al., 2005), Prdm14 (Yamaji et al., 2008) and Ap2γ (Weber et al., 2010). After E8.5, PGCs start migrating towards the genital ridges, and there, at E11.5, they undergo sex determination (Saga, 2008). In the gonads of female embryos, gonocytes are subjected to retinoic acid (RA), and thus, they enter meiosis (Koubova et al., 2006) (Figure 1.1 B). On the contrary, male gonads express high levels of Cyp26b1, which degrades RA and prevents male gonocytes from entering meiosis (Saga, 2008). Thus, male gonocytes undergo mitotic arrest at G1/G0 and remain quiescent until birth (McLaren, 2003; Saga, 2008).

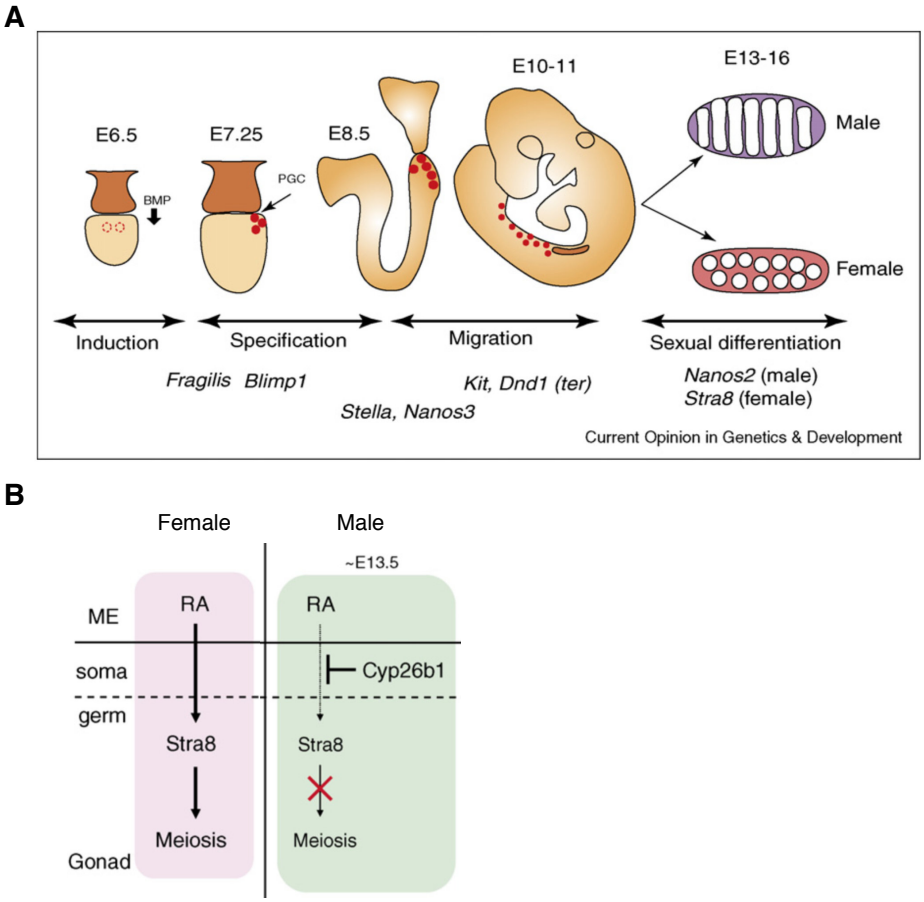


Figure 1.1. Germ cell development during mouse embryogenesis, adapted from (Saga, 2008).

A. The figure depicts the main stages of germ cell development between embryonic day (E) 6.5 and 16. Primordial germ cells (PGCs) are shown as red circles. Key genes involved in each phase of the process are also shown at the bottom. **B.** In female embryonic gonads (pink panel), retinoic acid (RA) induces Stra8 and triggers meiosis. In males (green panel), the meiotic entry of gonocytes is prevented by the expression of Cyp26b1.

After birth, the mitotically arrested male gonocytes re-enter the cell cycle and migrate from the central part of the seminiferous tubules towards their periphery. Once they reach the basement membrane of the tubules, at post-natal day P3-6, gonocytes become spermatogonia (Bellve et al., 1977). Spermatogonia are undifferentiated germ cells which undergo

spermatogenesis (Russell et al., 1990). However, the initial round of spermatogenesis after birth in mouse is initiated directly from gonocytes, thus bypassing the spermatogonia stage (Yoshida et al., 2006).

Spermatogenesis is a highly organised developmental process that results in the production of haploid spermatozoa, the male gametes competent for egg fertilisation (Phillips et al., 2010). This process takes place inside the seminiferous tubules of testes and involves different types of germ cells (Figure 1.2). These cell types arise in a sequential manner and are spatially arranged inside tubules (Russell et al., 1990). Male germ cells include the undifferentiated spermatogonia, which are initially located near the basal membrane of the tubules, but upon cell division and differentiation, they move towards the centre of the tubules, to the adluminal compartment (Figure 1.2) (Russell et al., 1990). Outside the seminiferous tubules, there is the interstitial tissue, composed of Leydig cells, which produce testosterone, blood vessels, and other somatic cells (Russell et al., 1990). However, inside tubules, there is only one type of somatic cells, the Sertoli cells. These are in close contact with germ cells, and they provide both mechanical and nutritional support, for each stage of spermatogenesis (França et al., 2016). Indeed, Sertoli are also called ‘nurse cells’, since they nourish germ cells by providing nutrients, such as lactate, cytokines, such as GDNF, and hormones (França et al., 2016). Moreover, adjacent Sertoli cells form tight junctions between each other, and this contributes to the formation of the blood-testis barrier (BTB), which separates two compartments within the tubules: the basal and the adluminal ones (Figure 1.2). This separation forms a unique microenvironment where, in the apical part, there are exclusively the meiotic germ cells and mature spermatozoa, which are completely isolated from the blood vessels (Jiang et al., 2014).

Importantly, almost all the functions performed by Sertoli cells are under the control of multiple signalling pathways (Ni et al., 2019).

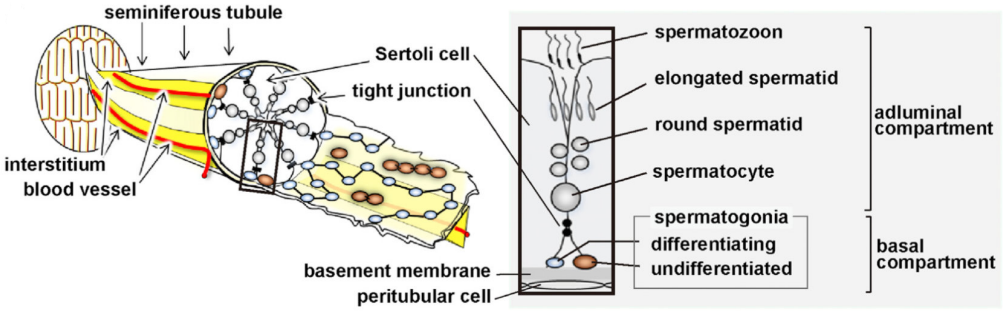


Figure 1.2. Seminiferous tubules showing the spatial arrangement of different types of germ cells, adapted from (Hara et al., 2014).

The figure shows the anatomy of seminiferous tubules, and the distribution of specific types of germ cells, in the basal compartment, and the adluminal or apical one. Undifferentiated spermatogonia are shown in brown and the other germ cells in grey.

As previously said, there are different types of germ cells, depending on their stage during spermatogenesis. In particular, spermatogonia were initially divided into three categories, type A, Intermediate (In) or type B (Russell et al., 1990) (Figure 1.3). This subdivision was based on the morphological features of their nuclei, which contained a different amount of heterochromatin, increasing from type A to type B (Phillips et al., 2010). Type A spermatogonia are undifferentiated cells, which also include the spermatogonial stem cells (SSCs) population (see below), and comprise A_{single} cells, which can divide to give rise to another identical A_{single} or generate two daughter cells called A_{paired}. These are the result of incomplete cytokinesis (Russell et al., 1990). A_{paired} can further divide and generate $A_{aligned}$ cells, which contain multiple cells in the same 'syncytia'. Long chains of $A_{aligned}$ spermatogonia undergo differentiation

and continue to proliferate to give rise to Intermediate spermatogonia, and then to type B spermatogonia (Figure 1.3). Then spermatocytes enter the meiotic divisions, to reduce their chromosome numbers and become haploid cells (Russell et al., 1990). Importantly, during meiosis, the genetic material is recombined through the crossing-over, in order to produce haploid spermatids with a unique genetic asset (Russell et al., 1990). Then, round spermatids undergo an additional developmental process in order to form specialised spermatozoa, which are released into the lumen of seminiferous tubules (Russell et al., 1990; Rooij, 2017).

The formation of long chains of germ cells implicates that cellular products are shared through intercellular bridges across the syncytia (Dunleavy et al., 2019). This peculiar feature is thought to contribute to the synchronisation of spermatogenesis, and also to make the haploid stages ‘phenotypically diploid’ (Haglund et al., 2011).

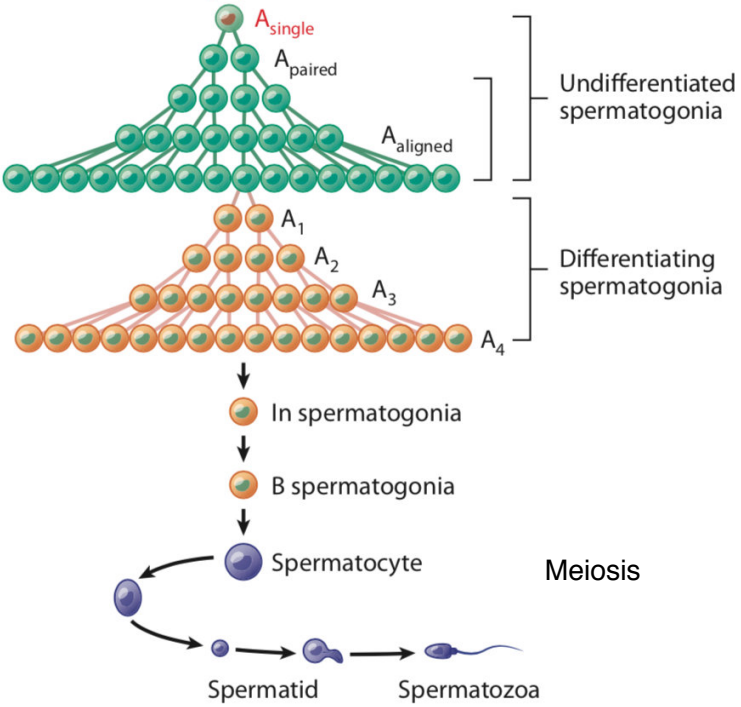


Figure 1.3 The different stages of mouse spermatogenesis, adapted from (Kanatsu-Shinohara and Shinohara, 2013).

Spermatogenesis starts from type A spermatogonia, which are divided into undifferentiated and differentiated cells. The undifferentiated population includes A_{single} spermatogonia, which can divide to produce an identical A_{single} or to generate A_{paired} spermatogonia through incomplete cytokinesis. Long chains of $A_{aligned}$ spermatogonia undergo multiple rounds of cell divisions and also differentiation. Intermediate spermatogonia (In) become type B spermatogonia and then spermatocytes, which enter meiosis. At the end of the process, spermatids undergo further maturation to produce spermatozoa.

The developmental processes which allow spermatogonia to become spermatocytes and then spermatids, not only take place in a spatially organised manner, but they are also specifically timed (Russell et al., 1990). Indeed, groups of germ cells present in a specific area of the tubule undergo

exactly the same sequence of events, every 8.6 days in mouse. This periodic process is called the spermatogenic cycle and includes twelve stages, each representing a unique combination of germ cells present in a distinct area of the tubule (Phillips et al., 2010; Rooij, 2017).

1.2 Spermatogonial stem cells (SSCs) and the models of self-renewal

Spermatogonial stem cells (SSCs) are unique as they can both self-renew to produce an identical SSC, and differentiate, to give rise to other spermatogonia, which undergo spermatogenesis (Kanatsu-Shinohara et al., 2008; Rooij, 2017)(Russell et al., 1990). Their self-renewal ability allows sustaining a long-term pool of SSCs, in the adult mouse, which fuels spermatogenesis throughout life. The first studies aimed at understanding SSCs division and self-renewal were based on the morphological features of spermatogonia, observed after fixation of seminiferous tubules (Chiarini-Garcia and Russell, 2001; Rooij and Russell, 2000). By evaluating their spatial arrangement within the tubules as well as the morphology of their nuclei, it was initially postulated that the most undifferentiated spermatogonia resided within the Type A and that only the A_s (A_{single}) were able to self-renew (Huckins, 1971; Oakberg, 1971). According to this model, later on called the 'A_s model' (Figure 1.4 A), an A_s can divide to form another identical A_s which also has self-renewal capacity, or it can divide to form two daughter cells, A_{pr} (A_{paired}), committed to differentiation (Huckins, 1971; Oakberg, 1971).

The identification of genes expressed only in selected groups of undifferentiated spermatogonia allowed defining better the properties of SSCs (Aponte et al., 2005). One of the first markers was the transcription factor neurogenin3, Ngn3, which was shown to be specifically expressed in undifferentiated spermatogonia (c-Kit-negative cells), and in A_s, A_{pr} A_{al4} ($A_{aligned4}$) cells (Yoshida et al., 2004). Moreover, the receptor for the glial cell line-derived neurotrophic factor (GDNF), called GFRα1, was shown to be expressed mainly in A_s (Hofmann et al., 2005), and to smaller extents also in

A_{pr} A_{al4} (Suzuki et al., 2009). Moreover, the RNA-binding protein NANOS2 was shown to be exclusively expressed in undifferentiated spermatogonia, and to nearly overlap with GFRα1 (Suzuki et al., 2009). The transcription factor PLZF was shown to be expressed in short chains of spermatogonia and to be critical for the maintenance of SSCs in mouse (Buaas et al., 2004; Costoya et al., 2004).

However, these genes were shown to be present simultaneously in A_s and in A_{pr}, as well as to a smaller extent in longer chains, which suggested the existence of heterogeneity within clusters of spermatogonia having the same morphological features (e.g., within A_s) (Suzuki et al., 2009). Later on, the inhibitor of differentiation 4, ID4, was shown to be present almost exclusively in A_s and to be necessary for the self-renewal ability of SSCs cultures (Oatley et al., 2011). Lineage tracing experiments also revealed that ID4-cells could give rise to long-term patches, which produce the entire spermatogenic lineage (Sun et al., 2015). Moreover, the self-renewal potential of ID4-cells was also tested with transplantation. This a functional assay which is based on the dissociation of donor testicular tissues and its transplantation into the empty testes of recipient mice, which are then examined for the ability to generate offspring (Brinster and Zimmermann, 1994). Thus, when transplanting ID4-cells into recipient mice, these were shown to repopulate testes, highlighting their self-renewal ability (Chan et al., 2014). However, not all A_s in testes express ID4, and within the ID4-cells, there is an expression gradient which correlates with their self-renewing ability (Helsel et al., 2017). Overall, these data suggested a hierarchy within the A_s, where only some rare A_s have the highest self-renewal ability and thus these were called 'the ultimate' SSCs (Helsel et al., 2017; Rooij, 2017). Since low levels of ID4 indicate gradual loss of self-renewal potential, these cells have been called 'transitory' SSCs (Helsel et al., 2017). According to this model, called the 'hierarchical A_s model' (Figure 1.4 B), upon A_s division, ID4 levels decrease and these daughter cells have higher chance to form A_{pr}. Moreover, cells expressing high levels of ID4 also

highly express BMI1, which marks a population of A_s with similar properties to ID4 (Komai et al., 2015).

However, live imaging and pulse labelling experiments on the GFRα1- and Ngn3-positive spermatogonia supported a different model for SSCs self-renewal, which was called the 'fragmentation model' (Hara et al., 2014; Nakagawa et al., 2010; Rooij, 2017)(Figure 1.4 C, D). According to this model, also A_{pr} and A_{al} cells can contribute to the self-renewing pool, by fragmenting and forming new A_s cells and smaller chains (Hara et al., 2014). Previous studies showed that Nng3 expression was restricted to the undifferentiated spermatogonia (Yoshida et al., 2004, 2006). Moreover, a pulse-labelling system which studied the behavior of Ngn3-cells, revealed the presence of two stem cell populations, the actual and the potential stem cells (Nakagawa et al., 2007). Potential stem cells normally have a rapid turnover, but upon loss of the actual stem cells, they are thought to switch towards self-renewal, suggesting a flexible behavior of spermatogonia, in contrast with the 'hierarchical A_s model' (Nakagawa et al., 2007). These studies were supported by further data afterwards. In the steady-state, GFRα1-cells, which are mainly represented by A_s, A_{pr}, and A_{al} cells, are at the top of the SSCs hierarchy, as demonstrated by live imaging and lineage analyses (Nakagawa et al., 2010; Sada et al., 2009). GFRα1-cells (in pink, Figure 1.4 D) normally maintain their own population and they also generate the second one, marked by Ngn3 (in green, Figure 1.4 D), which encompasses more A_{al} cells. Ngn3-cells mainly give rise to the differentiating c-Kit-cells (in blue, Figure 1.4 D), but they also have the ability to fragment into shorter chains and even into A_s (Nakagawa et al., 2010). These fragmented cells also re-acquire the expression of GFRα1 (yellow arrows, Figure 1.4 D) and can contribute to the SSCs pool in the long-term. Moreover, during regeneration, the reversibility of the Ngn3 population becomes more accentuated, suggesting that the Ngn3-cells may act as a reserve for the actual stem cells, following tissue insult (Hara et al., 2014; Nakagawa et al., 2007, 2010).

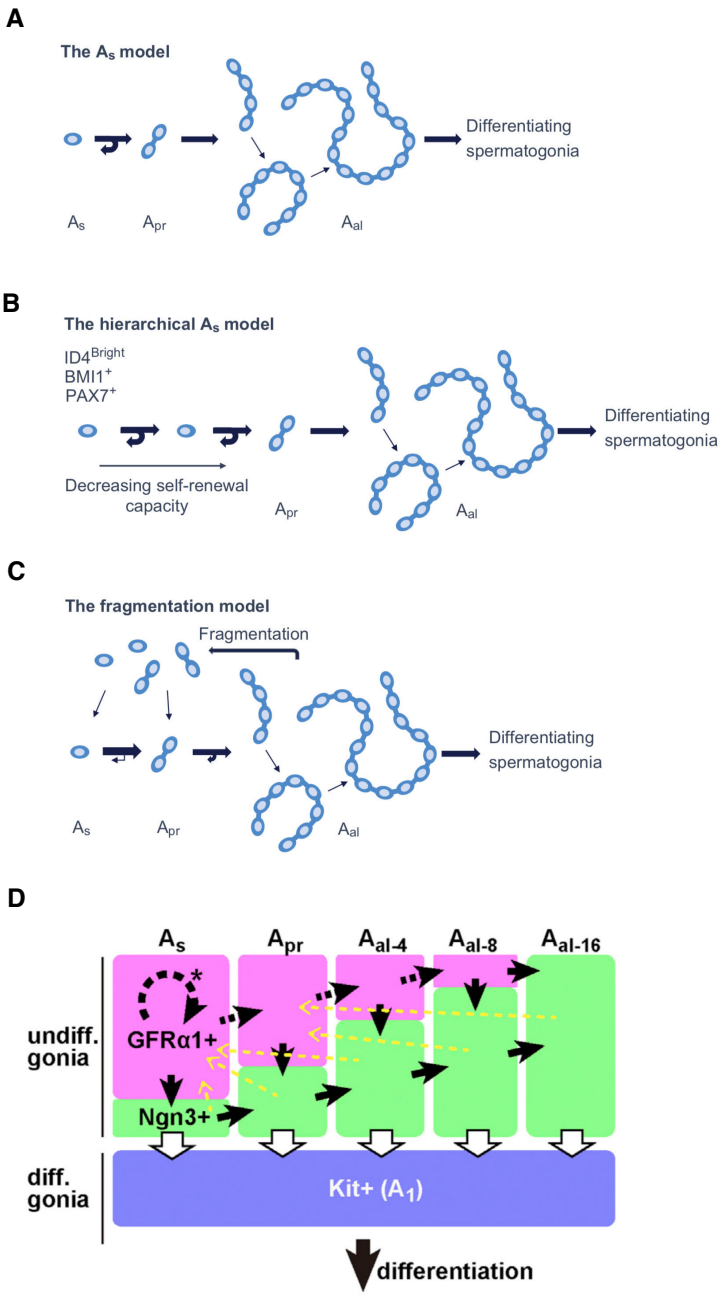


Figure 1.4 Models for SSCs self-renewal and division, adapted from (Hara et al., 2014; Rooij, 2017).

A. According to the historical 'A_s model', the A_s spermatogonia were the only stem cells. Upon division, A_s could either self-renew or produce two daughter cells (A_{pr}), which were already committed towards differentiation and could not contribute to the

self-renewing pool. **B.** In the 'A_s hierarchical model', the A_s population is heterogeneous, as it is marked by different levels of ID4 and BMI1. Their expression levels correlate with their self-renewal ability, as the ID4-bright cells have the highest self-renewal potential. **C.** In the 'fragmentation model', most of the times A_s generate $A_{pr,}$ which then form A_{al} cells. Chains of A_{pr} and A_{al} can also revert to A_s, by fragmentation, thus, implicating that A_{pr} and A_{al} cells are not completely committed for differentiation. **D.** Hierarchy among different subpopulations of spermatogonia, according to the fragmentation model. Undifferentiated spermatogonia are marked by GFRa1 (in pink) and Ngn3 (in green), whereas differentiating spermatogonia are c-Kit+ (blue). In this model, GFRa1+ A_s, marked by the asterisk (*), are the self-renewing pool; the reversibility of Ngn+ spermatogonia is depicted by yellow arrows.

1.3 SSCs cultures *in vitro*

SSCs constitute a very tiny population *in vivo*, as they are estimated as 0.02-0.03% of all germ cells in the mouse testes (Rooij and Russell, 2000). Their scarcity has hampered for decades the study of the molecular processes underlying their self-renewal ability. However, after identifying the cytokine GDNF as a critical regulator of SSCs self-renewal (Meng et al., 2000), scientists developed the first SSCs culture system, which allowed the exponential expansion of SSCs *in vitro*, and their long-term culture (Kanatsu-Shinohara et al., 2003). SSCs were originally derived from neo-natal mice, at P0-2, a period during which testes are enriched for SSCs (Rooij, 2017). These cells were grown on mouse feeders (MEFs) and formed grape-like colonies, in the presence of four cytokines GDNF, FG2, EGF and LIF, essential for SSCs self-renewal (Kubota et al., 2004). Importantly, upon transplantation into recipient testes, SSCs from these cultures were able to start spermatogenesis and generate fertile offspring (Kanatsu-Shinohara, 2005). However, only 1-2% of these SSCs were shown to produce colonies *in vivo*, highlighting the functional heterogeneity of these cells cultures (Kanatsu-Shinohara and Shinohara, 2013). Another important property of SSCs cultures is to maintain

in the long term a stable karyotype and a correct DNA methylation pattern in imprinted genes, even after two years of culture (Kanatsu-Shinohara, 2005). Later on, SSCs cultures were also developed in feeders-free conditions as well as in serum-free cell media (Kanatsu-Shinohara et al., 2005, 2014).

Moreover, subsequent studies showed that SSCs could also be derived from neonatal testes at P4.5-7.5 (Kubota et al., 2004), and even from 4-6 weeks old mice (Guan et al., 2006). Although germ cells are thought to lose their pluripotency during early embryonic development (Takashima and Shinohara, 2018), it was shown that SSCs derived from neonatal testes have the ability to spontaneously convert into pluripotent cells *in vitro* (Kanatsu-Shinohara et al., 2004). Additionally, also SSCs derived from adult testes were shown to retain pluripotency, when cultured under specific conditions (Guan et al., 2006).

As previously mentioned, SSCs can divide logarithmically for a long time in culture, in the presence of four cytokines, among which GDNF (Takashima and Shinohara, 2018). Notably, GDNF mediates the activation of the phosphatidylinositol 3-kinase-AKT pathway (Figure 1.5), which is necessary for SSCs self-renewal, but when this pathway is constitutively activated, also independently of GDNF, SSCs can still proliferate (Lee et al., 2007; Oatley et al., 2007). Both GDNF and FGF2 stimulate the Src kinases, which, in turn, activate Ras and subsequently, the Akt and MEK pathways (Takashima and Shinohara, 2018). In particular, activation of the MAP2K mediated by FGF2 upregulates ETV5 and BCL6B, which then stimulate self-renewal (Ishii et al., 2012; Oatley et al., 2006). Since ETV5 controls the expression of c-ret (Tyagi et al., 2009), a component of the GDNF receptor essential for spermatogenesis (Naughton et al., 2006), these pathways also use feedback mechanisms to support SSCs self-renewal.

Moreover, other studies discovered that self-renewal is also promoted by moderate levels of ROS (reactive oxygen species) produced by two NAPDH oxidase, Nox1 and Nox3, which are also positively regulated by the AKT and MAP2K1 pathways (Morimoto et al., 2013, 2015).

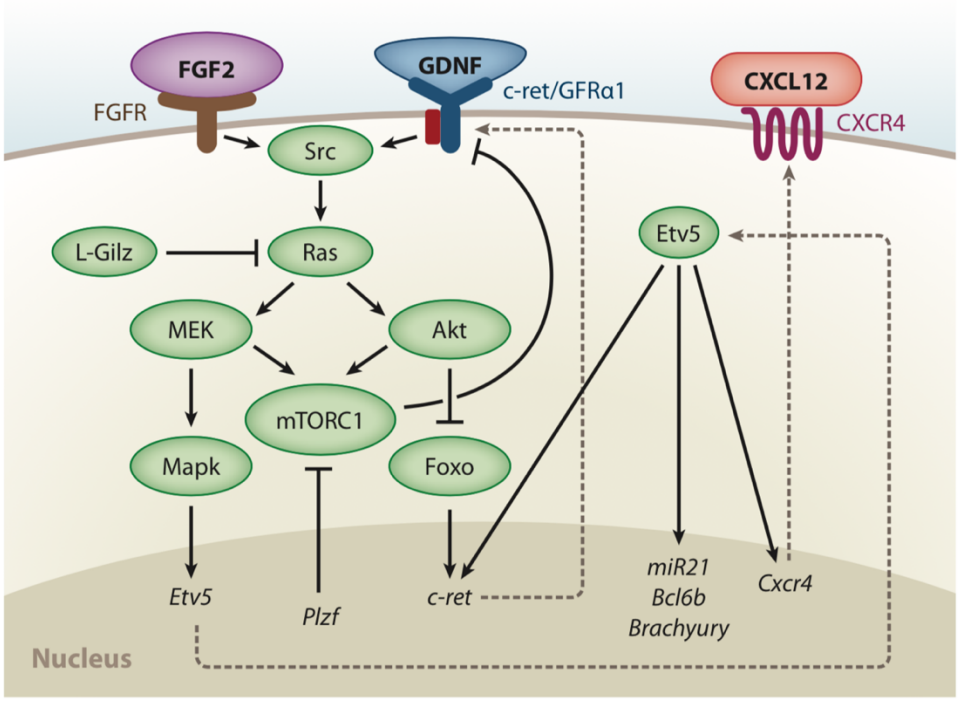


Figure 1.5. The main signalling pathways involved in SSC self-renewal, adapted from (Kanatsu-Shinohara and Shinohara, 2013).
On SSCs membrane, GDNF and FGF2 bind to the FGFR and GFRa1/c-ret receptors respectively and trigger signalling cascades which sustain SSC self-renewal.

1.4 Post-transcriptional gene regulation (PTGR)

Post-transcriptional gene regulation (PTGR) is an efficient way to modify the level of transcripts in response to cellular or extra-cellular stimuli. In contrast with the transcriptional regulation of gene expression, PTGR acts rapidly since it uses existing RNAs, which are already transcribed. This type of regulation occurs in every type of mRNA although to different extents, and takes place also as a quality control process, from the very early phases of an mRNA life until its end.

During transcription, nascent mRNAs are processed, to ensure their correct maturation. These co-transcriptional events include the addition of the 5'-cap to form a 7-methylguanosine, the splicing, and the 3'-end maturation, which includes cleavage and polyadenylation. If any of these early events do not occur, mRNAs are degraded, without even being exported from the nucleus to the cytoplasm (Corbett, 2018; Singh et al., 2015). Following nuclear export, mRNAs are subjected to additional quality-control processes, to ensure that only 'healthy' mRNAs are translated, thus optimising cellular energy. For instance, the Non-sense mediated decay (NMD) pathway induces the degradation of mRNAs with premature stop codons, thus preventing the production of aberrant proteins (Lykke-Andersen and Jensen, 2015).

From the birth to the end of their life, mRNAs are accompanied by other factors which assemble along their body length. It is estimated that on average, 14 RNA-binding proteins (RBPs) bind to the 3' UTR of each mRNA (Plass et al., 2017). The combination of mRNAs, RBPs as well as other factors, such as miRNAs, form the messenger ribonucleoprotein particles (mRNPs). mRNPs can have variable composition depending on the specific phase of the mRNA lifecycle or simply on the localisation of the mRNA. The exact array of factors included in a specific mRNP decides the fate of a given mRNA, its expression level and in general they are the major determinant of PTGR (Kuersten and Goodwin, 2003).

RBPs in mRNPs can be either general or specific. Examples of general RBPs which bind to regions shared among all mRNAs, are: the nuclear cap-binding complex, CBC20/80, which binds to the 5' cap of mRNAs, the Exon junction complex, (EJC), which binds to the exon-exon junctions, and the Cytoplasmic poly(A) binding protein, (PABPC), which binds to the poly(A) tail of every mRNA (Moore, 2005).

RBPs can associate along the entire length of the mRNA body without sequence specificity, such as the Y-box proteins, which are useful for the packaging of mRNPs. However, RBPs can also bind with sequence specificity to certain regions of mRNAs, enclosed in the 5' and 3' UTRs (Moore, 2005).

For instance, 3' UTRs with AU-rich elements (ARE) are bound by the TTP (tristetraprolin) proteins, which trigger RNA degradation (Carballo et al., 1998a). In other cases, mRNAs have specific miRNA-recognition elements, which are bound by miRNAs. These, in turn, associate with the RBPs Argonaute and induce mRNA repression (Mayya and Duchaine, 2019).

Therefore, the nucleotide sequence of mRNAs is critical to determine the fate of transcripts.

Moreover, the promoter sequence of a gene can also affect the decay of the corresponding transcripts (Haimovich et al., 2013). Since this strategy is adopted by yeast and mammalian cells, it suggests that coordination between transcription and mRNA decay is a highly conserved process which might also have helped the evolution of living organisms (Haimovich et al., 2013).

Another element that affects PTGR is the 3' UTR structure, which has been recently defined as a new level of genetic code (Bevilacqua et al., 2016; Mayya and Duchaine, 2019). In contrast to DNA, single-stranded RNA can fold into a myriad of secondary and tertiary structures, which also contribute to the variety of RNA functions. Secondary structures are based on canonical (G-C and A-U) and noncanonical base pairs, such as the G · U wobble, but also include unpaired nucleotides. Altogether, these give rise to hairpins, which are composed of stem and loops, bulges, and 3-way junctions. Then, secondary structures fold into complex tertiary structures, which can include pseudoknots and G-quadruplexes.

This knowledge is used to predict the RNA structure with free energy minimization, in traditional, thermodynamics-based approaches. However, these *in silico* predictions are not sufficient to grab the complexity of RNA structures in cells. RNA structure is also affected by RBPs and other factors (like lncRNAs and miRNAs) existing *in vivo*, which modify the accessibility of certain regions in the mRNA. Thus, to capture the complexity of the RNA structure in cellular systems, existiting methods have been recently coupled to next-generation sequencing and applied to various organisms. For instance, new insight into the ‘RNA structurome’ concerns stress-related mRNAs. These

were shown to have more single-stranded regions than previously predicted with traditional methods. Notably, these mRNAs are thought to be very dynamic *in vivo*, a property that could confer them the necessary flexibility to respond to stress and to adapt to other regulatory processes in cells (Bevilacqua et al., 2016; Jacobs et al., 2012; Kwok et al., 2015).

1.5 The role of RNA-binding proteins (RBPs) in PTGR

RBPs greatly participate in the PTGR. They associate with mRNAs throughout their lifecycle and thus are major determinants of mRNA fate. Importantly, also the 'history' of RBPs bound to a specific transcript can greatly influence its fate. One example is represented by the EJCs which are deposited on mRNAs in the nucleus. Although this deposition occurs at the beginning of mRNA life, it can have downstream effects in the cytoplasm, when, in some cases, it triggers the NMD pathway and determine RNA degradation (Dreyfuss et al., 2002).

RBPs are widely employed by cells, in all organisms studied. RBPs are remarkably conserved across evolution, in particular, compared to TFs. Notably, 50% of human RBPs are also found in *S. cerevisiae*, whereas only 14% of human TFs are also found in *S. cerevisiae*. This highlights the fact that RBPs diversified very early in evolution compared to TFs. Thus, if TFs expansion reflects better the complexity of organisms during evolution, RBPs indicate highly conserved metabolic processes shared in all living organisms (Gerstberger et al., 2014).

On average, human tissues express RBPs at higher levels than any other category of proteins. Moreover, although similar numbers of genes encode RBPs and TFs, RBPs constitute 20% of protein-coding transcripts and TFs, only 3% (Gerstberger et al., 2014).

Furthermore, although the majority of human RBPs has no tissue-specificity, 90% of the tissue-specific RBPs reside in the germline, brain, muscle, bone marrow, and liver cells (Gerstberger et al., 2014). Moreover, ~50% of these

specialised RBPs is enriched in adult testis (Gerstberger et al., 2014). This is not surprising, given that RBPs have essential roles during spermatogenesis, where they are involved in splicing, in the 3' end formation of mRNAs, in translation, and RNA decay (Elliott, 2003).

RBPs function greatly depends on their ability to bind and select their RNA targets. The interaction between proteins and RNAs is mediated by RNA-binding domains (RBDs), which include RRM, KH, and zinc fingers, among others. In general, target specificity is achieved by the particular features of each domain (described below), but also by the presence of multiple repeats of the same domain or by the combination of different domains. Overall, multiple domains increase the sequence specificity of an RBP, although RNA-binding can also be greatly influenced by either synergy or competition among different RBPs (Lunde et al., 2007).

The most common module used by RBPs is the RNA recognition motif (RRM). This usually contacts ~4-8 nucleotides of RNA, through an Arg or Lys and two aromatic amino acids present on the surface of a beta-sheet (Lunde et al., 2007). The hnRNP K homology (KH) can bind to RNA by recognising ~4 nucleotides, through hydrogen bonds and electrostatic interactions formed in a cleft of the KH domain. Zinc fingers include several types of domains, which are classified depending on the number and order of cysteine (C) and histidine (H) residues which coordinate zinc. Thus, there are C2H2, CCCH, and CCHC domains, which use slightly different ways to achieve target recognition.
Other domains include the S1 domain, which binds to RNA similarly to the RRM, and the PAZ and PIWI domains which are specifically present in small non-coding RNA-binding proteins (Lunde et al., 2007). RBPs can also have double-stranded RNA-binding domains (dsRBD) such as the RNA editing enzyme ADAR as well as DICER (Saunders and Barber, 2003). However, the dsRBDs generally bind to RNA without sequence specificity, by contacting the phosphate backbone and the 2'-OH groups of RNAs.

Interestingly, it has been recently discovered that also many proteins with unconventional domains can bind to RNA. Notably, these domains include intrinsically disordered regions (IDRs), the enzymatic cores of metabolic proteins, as well as protein-protein interfaces (Dreyfuss et al., 2002; Hentze et al., 2018).

1.6 Cytoplasmic mRNA degradation

Cells do not accumulate as much RNA as they transcribe it, indicating that they actively employ RNA degradation pathways.
By default, mRNAs are unstable molecules, prone to decay, whose life is constantly threatened by the action of RNases, ubiquitously present in cells. However, their instability also makes them an ideal means to control gene expression.
Indeed, mRNAs are degraded for various reasons in cells. First of all, to prevent their permanent expression throughout time, which would promote the wrong cellular program. Then, mRNA decay is used to respond to changes in the microenvironment rapidly. Moreover, it is also employed to perform quality control, as defective mRNAs are promptly removed in cells, to prevent the production of aberrant proteins.
In general, an abnormal accumulation of RNA in cells would be detrimental, because free and random RNAs molecules could sequester RBPs, and impair their functions. Furthermore, free nuclear RNA fragments could form RNA-DNA hybrids, which would negatively impact DNA replication (Houseley and Tollervey, 2009).

mRNA decay is mediated mainly by exonucleases, which degrade RNAs from the two extremities, the 5' and 3' ends. This is the main reason why every mRNA has to be protected by the 7-methylguanosine cap at the 5', and by the poly(A) tail at the 3' end.

The prevalence of exonucleases, rather than endonucleases, in RNA decay pathways, seems to have evolved in order to protect cells from an excess of free RNA fragments. Since small noncoding RNAs are extensively used to regulate gene expression, increased production of random small RNAs in cells could cause their inappropriate entry into the RNA processing pathways (Bühler et al., 2008; Houseley and Tollervey, 2009).

Since every mRNA is capped and polyadenylated, and the removal of these two terminal structures is considered the most rate-limiting step in RNA decay (Meyer et al., 2004).

In general, mRNA degradation is initiated by the shortening of the poly(A) tail catalysed by the conserved PAN2/3 and CCR4-NOT (CNOT) complexes. This occurs via a 'biphasic model' where PAN2/3 start to trim the poly(A) tail, and then CNOT completes the deadenylation, with its two catalytic subunits, CCR4 and CAF1 (Figure 1.6 B) (Yamashita et al., 2005).

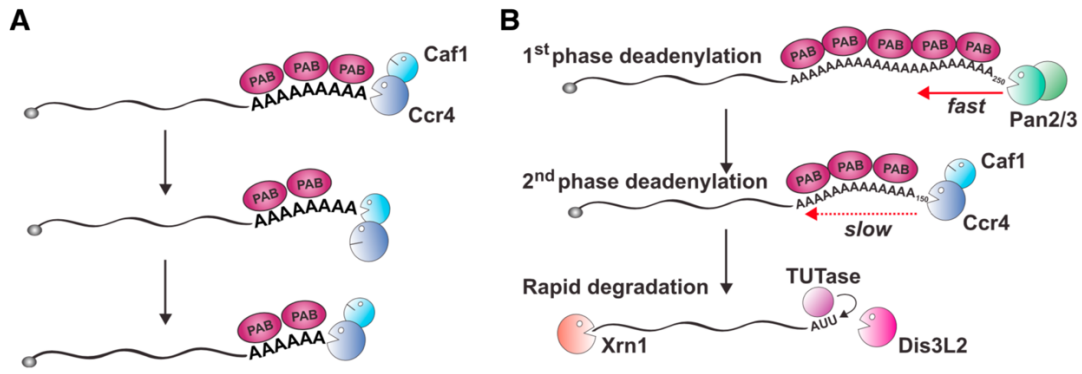


Figure 1.6. Models for the cytoplasmic mRNA degradation (Bresson and Tollervey, 2018).

A. CAF1 (CNOT7/8 in mammals) and CCR4 (CNOT6/6L in mammals) alternate on the poly(A) tail of mRNAs. CCR4 preferentially binds to tails covered by PAB, which stimulates its deadenylation activity, whereas CAF1 prefers trimming free poly(A) tails. **B.** In the first phase of deadenylation, PAN2/3 deadenylate long polyA tails (> 150 nt), rapidly. Then, once tails become shorter than 150 nt, CAF1 and CCR4

replace PAN2/3 and catalyse a slower deadenylation. mRNAs with short poly(A) tails are degraded from the 5' end by XRN1, after removal of the 5' end cap. In some cases, very short tails are targeted for uridylation by the TUT enzymes, which then recruit DIS3L2 to complete the degradation, (3'-5' direction).

After deadenylation, short poly(A) tails are degraded from the 5' end, by the action of the Xrn1 exonuclease, which acts after decapping (Bresson and Tollervey, 2018; Wahle and Winkler, 2013).

More recently, two studies in yeast and human cells have enlarged our knowledge about the 'biphasic model' and the relationship between the cytoplasmic deadenylases and the poly(A) binding proteins, called PAB1 in yeast, and PABPC1 in human (Webster et al., 2018; Yi et al., 2018).

In yeast, PAB1 was shown to physically interact with CNOT and to accelerate the deadenylation process. In particular, PAB1 stimulates CCR4 and inhibits CAF1. This would cause the alternation of CCR4 and CAF1 on the poly(A) tails, with CCR4 binding to tails still covered by the PAB1, and CAF1 binding only to free tails (Figure 1.6 A) (Webster et al., 2018). Moreover, by measuring the length of the poly(A) tail by TAIL-seq, Yi and colleagues discovered that deadenylation occurs as discrete steps, by removing 27 adenine residues in each cycle, a length that corresponds to the PABPC1 footprint (Yi et al., 2018). Additionally, the first phase of deadenylation, mediated by PAN2/3, occurs very rapidly, on tails longer than 150 nt. In contrast, the second phase, mediated by CNOT, is slower and occurs on tails shorter than 150 nt (Figure 1.6 B).

Importantly, in both yeast and human, PAB1 and PABPC1 activate deadenylation. Notably, PABPC1 promotes mRNA decay exclusively through CCR4, thereby protecting mRNAs from over-deadenylation. Thus, although on one hand PABPC1 coats poly(A) tails and increases translation efficiency of mRNAs, on the other hand, it also stimulates mRNA decay (Bresson and Tollervey, 2018).

mRNAs degradation is also regulated by *cis*-elements, present in the 3' UTR. The most characterised and widespread sequences are the AU-rich elements (AREs) (Otsuka et al., 2019). These are constituted by the nonamer 'UAAAUAUU' (Worthington et al., 2002) and are recognised by several RBPs, which can have both stabilising and destabilising functions. For instance, upon binding to ARE-mRNAs, TTPs (Carballo et al., 1998b) and BRF1 (Stoecklin et al., 2002) decrease their half-life through the recruitment of other proteins, such as CNOT, involved in the RNA decay pathways (Mayya and Duchaine, 2019). However, when HuR proteins bind to AU-rich sequences, they can elicit opposing effect, by impairing mRNAs deadenylation and thus slowing their decay (Peng et al., 1998).

Thus, only a complex combination of multiple factors contributes to the fine regulation of ARE-mRNAs. ARE-RBPs can compete with each other by binding to closely related ARE sequences. Moreover, ARE-RBPs are post-translationally modified, and this constitutes an additional layer of regulation. Eventually, RNA secondary structures in the proximity of the ARE elements as well as other unknown *cis*-elements in the 3' UTRs, could contribute to the final mRNA stability of ARE-mRNAs (Mayya and Duchaine, 2019; Otsuka et al., 2019).

Apart from the ARE elements, which are an integral part of some mRNAs and reflect the original DNA sequence, also other types of elements present in the poly(A) tail can affect mRNA stability. These modifications are called 'RNA tailing' and are the result of non-templated nucleotides addition mediated by different enzymes, at the posttranscriptional level (Lee et al., 2014; Norbury, 2013). In particular, the oligo-uridylation of short poly(A) tails (< 25 nt) mediated by the TUT proteins stimulates mRNA decay (Chang et al., 2014; Rissland and Norbury, 2009; Lim et al., 2014; Morgan et al., 2017). Moreover, mRNAs longer than 25 nt can also be guanylated. Guanosine is added by the TENT4A and TENT4B enzymes as single residues, at the very last position of

the poly(A) tails or penultimate positions. These modifications have been shown to protect mRNAs from deadenylation mediated by the CNOT complex (Lim et al., 2018).

In summary, *cis*-elements in the 3' UTR, specific RBPs, and RNA tailing explain the complex mechanisms underlying mRNA instability, and the variability of mRNAs half-life which, in mammals, ranges from 15 minutes to 24 hours (Meyer et al., 2004).

1.7 Structure and functions of the CCR4-NOT (CNOT) complex

The CCR4-NOT complex (CNOT) is the most important cytoplasmic deadenylase, in all the organisms analysed so far. Although this complex is also known to have additional functions in the nucleus related to transcription and nuclear surveillance (Inada and Makino, 2014), I will focus on the cytoplasmic roles of its subunits and in particular on deadenylation.
The complex is composed of multiple subunits with different structures and functions, and only some of them have catalytic activity (Figure 1.7).

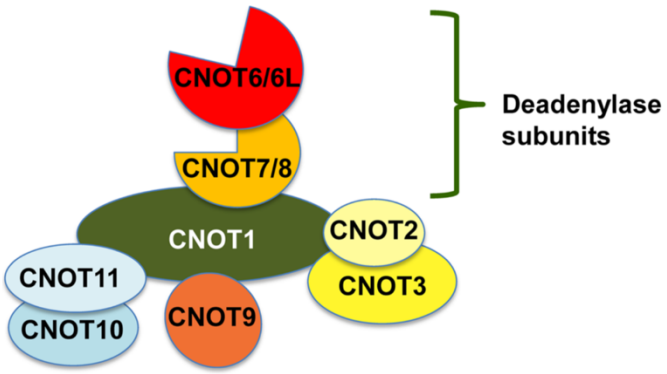


Figure 1.7 Components of the CCR4-NOT complex, from (Shirai et al., 2014).
The CCR4-NOT complex is the major cytoplasmic deadenylase and is composed of many subunits. CNOT1 is the largest one and serves as a scaffold for all the

remaining components. CNOT7/8 (CAF1 in yeast) and CNOT6/6L (CCR4 in yeast) are the only subunits with deadenylase activity, which results in the removal of the poly(A) tails from mRNAs. Importantly, CNOT6 and CNOT6L do not directly interact with CNOT1, but they are recruited by CNOT7/8 into the CCR4-NOT complex.

The NOT1 subunit is the largest one and serves as a scaffold to assemble all the other components of the complex. Importantly, the middle region of NOT1 directly interacts with CNOT7 or CNOT8 (CAF1 in yeast), which are two catalytic components. Additionally, the same part of the protein also associates with CNOT9. The N-terminal part of CNOT1 interacts with CNOT11 which, in turn, recruits CNOT10 into the complex. The C-terminal region interacts with CNOT2 and CNOT3 (Bawankar et al., 2013; Shirai et al., 2014). Although CNOT1 has no catalytic domain, it is essential for the deadenylase activity of the entire complex and for its correct assembly (Ito et al., 2011). Importantly, CNOT1 also mediates protein-protein interactions with RBPs, which recruit CNOT to specific mRNAs (Shirai et al., 2014).

CNOT2 and CNOT3 are two structural components of the CNOT complex, as they do not have any catalytic activity. They share a NOT-box domain in their C-terminal region (Zwartjes et al., 2004) and CNOT2 serves as a molecular bridge between CNOT1 and CNOT3. Both subunits positively regulate the deadenylation activity of the complex and, importantly, *Cnot3*$^{-/-}$ mice are embryonic lethal (Neely et al., 2010; Shirai et al., 2014).

CNOT9 is another structural subunit of the complex, which interacts with CNOT1 with the ARM (armadillo) repeat domain. Interestingly, the same domain is also able to bind to ssDNA polymers, apart from poly(dA). CNOT9 is expressed at very low levels in human tissues, apart from testis and in mouse it is expressed in various organs, including testis (Chen et al., 2001; Garces et al., 2007).

CNOT4 is an E3 ubiquitin ligase, which is not constitutively present in the human CNOT complex and indeed is not essential for its deadenylase function. However, in yeast, it is involved in the co-translational degradation of aberrant proteins stalled at the ribosomes (Collart and Panasenko, 2012). Recently, it has been shown that the C-terminus of CNOT4 contains a conserved motif, both in *Drosophila* and human. This region binds to the CNOT9 subunit in order to recruit CNOT4 into the CNOT complex (Keskeny et al., 2019).

CCR4 and CAF1/POP2 are the catalytic subunits of the CNOT complex (Wahle and Winkler, 2013). In mammals, the yeast CCR4 has two orthologs, CNOT6 and CNOT6L. All the CCR4 orthologs share two evolutionary conserved domains: the exonuclease/endonuclease/phosphates (EEP) domain, located in the C-terminal region, and the leucine-rich repeat (LRR) domain in the N-terminus. The EEP domain has deadenylase activity, whereas the LRR allows CCR4 to physically interact with CAF1 (Winkler and Balacco, 2013). The crystal structure of the human CNOT6L showed that this protein exclusively binds to poly(A) substrates and that it prefers RNA to DNA (Wang et al., 2010). In human cells, CNOT6 and CNOT6L do not co-exist in the same CNOT complex, suggesting that they might be redundant (Lau et al., 2009). Interestingly, it was recently shown that the genetic deletion of *Cnot6l* causes female infertility, due to accumulation of the maternal mRNAs, in mouse oocytes (Sha et al., 2018).

The other catalytic subunit of the CNOT complex is CAF1 (in yeast), which has two orthologs in mammals, CNOT7 and CNOT8. These proteins belong to the conserved class of DEDD exonucleases, as they share the ribonuclease D domain. This domain spans most of the length of the proteins (Collart and Panasenko, 2012; Shirai et al., 2014; Winkler and Balacco, 2013). CNOT7 and CNOT8 directly bind to CNOT6 and are essential for the recruitment of

CNOT6/6L to the CNOT complex. Importantly, Cnot7 null mice display severe impairment of their fertility (Berthet et al., 2004; Nakamura et al., 2004).

1.8 The evolutionarily conserved family of RNA-binding proteins NANOS

1.8.1 *Nanos* genes family

Nanos genes are characterised by a highly conserved sequence which encodes two CCHC zinc finger motifs, usually positioned at the C-terminus of NANOS proteins (Figure 1.8) (e.g., in *Drosophila* NANOS is 401 aa long, and the zinc fingers (ZFs) are included between 318-372 aa; in mouse, NANOS2 is 136 aa long, and its ZFs are between 60-114 aa; in human NANOS1 is 292 aa long, and its ZFs are positioned between 213-267 aa). Another conserved region present in most of *Nanos* genes encodes the NOT1 interacting motif (NIM) at NANOS N-terminus. Although the NIM region is shared among all vertebrates and some invertebrates, it is not as conserved as the zinc finger domain. Thus, the Zinc fingers have been used to draw a *nanos* phylogenetic tree, which showed that *nanos* genes are present in all animals (De Keuckelaere et al., 2018). These include the ancestral Non-bilaterian animals, like Ctenophora, Porifera and Cnidaria as well as Protostomes, with *Drosophila* and *C. elegans* among others.

Moreover, *nanos* genes are also present in Deuterostomes including Fish (e.g. *Danio rerio*), Amphibia (*Xenopus tropicalis*), Rodents and Primates. The number of *nanos* genes per animal can vary from one to four. Only one gene is present in *Drosophila*, whereas two genes are found in Cnidaria (e.g. *Hydra vulgaris*), and three genes are present in *C. elegans* and most of the vertebrates, including mouse and human. Eventually, four genes can be found rarely, as in Fish.

Interestingly, among vertebrates, NANOS paralogs (like NANOS1, NANOS2, and NANOS3) are not as conserved as NANOS orthologs, found in different

species. This suggests that the original *nanos* gene underwent duplication events, which generated paralogs with probable distinct functions.

Moreover, although the zinc finger domain is highly conserved, the sequence encoding for it is highly variable in length. For instance, in *Drosophila*, the only *nanos* gene present encodes a large protein of 401 amino acids (aa). In contrast, mouse and human *nanos* genes produce small proteins, such as NANOS2 (136 and 138 aa respectively) or bigger ones like NANOS1 (267 and 292 aa respectively) (De Keuckelaere et al., 2018).

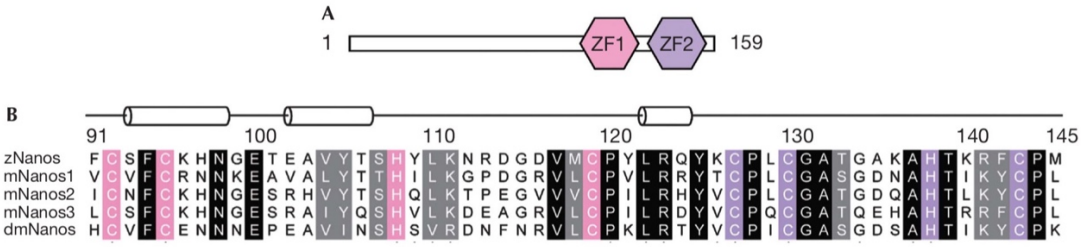


Figure 1.8. Schematic representation of the Nanos proteins, adapted from (Hashimoto et al., 2010b).

A. The figure shows the architecture of zebrafish Nanos protein, which is representative of Nanos orthologs. In particular, in their C-terminal region, there are two conserved zinc fingers, ZF1 and ZF2. **B.** The figure shows the amino acid sequence alignment of the C-terminal portion of different Nanos proteins. zNanos is zebrafish Nanos, mNanos1, -2, -3 are mouse Nanos, dmNanos is *Drosophila* Nanos. Amino acid residues in pink are present in the ZF1, whereas residues in violet are the ones in ZF2.

1.8.2 Functions of *Nanos* genes

NANOS proteins share the evolutionary conserved zinc fingers domain as well as similar functions in germ cell development.

In particular, this role is conserved in Protostomes, like *Drosophila* (Kobayashi et al., 1996; Lehmann and Nusslein-Volhard, 1991), and *C. elegans* (Subramaniam and Seydoux, 1999), and, among Vertebrates, in *Xenopus* (MacArthur et al., 1999), zebrafish (Köprunner et al., 2001), and mouse (Tsuda et al., 2003).

In *Drosophila*, *nanos* was shown to be essential for the migration of pole cells towards the gonads, and the formation of functional germ cells. Upon *nanos* loss, pole cells activate prematurely the germline expression program, that usually is switched on only after gonad colonisation (Forbes and Lehmann; Kobayashi et al., 1996).

Notably, Nanos and Pumilio are involved in the mitotic quiescence of pole cells, by binding to a *nanos* response element (NRE) (Wharton and Struhl, 1991) in the 3' UTR of *cyclin* B mRNA (Asaoka-Taguchi et al., 1999; Dalby and Glover, 1993). First, Pumilio recruits Nanos to the NRE, and then Nanos recruits the entire CCR4-NOT complex to *cyclin* B. This is achieved through a direct interaction between Nanos and the NOT4 subunit, which eventually stimulates the translational repression of *cyclin* B (Kadyrova et al., 2007).

Nanos regulates pole cells also by suppressing apoptosis (Hayashi et al., 2004). Similar to the repression of *cyclin* B in the migrating germ cells, Nanos also inhibits the pro-apoptotic *hid* mRNA at the translational level (Sato et al., 2007).

Additionally*, Nanos* was shown to be involved in germline self-renewal. In *nanos* mutants, there is a loss of germline stem cells (GSCs), which leads to empty ovaries (Wang and Lin, 2004). In particular, in the absence of the Nanos-Pumilio complexes, GSCs do not self-renew because there is a de-repression of the germline differentiation program (Gilboa and Lehmann, 2004; Wang and Lin, 2004). This process is thought to occur by controlling *mei-P26* mRNA. Again, the Nanos-Pumilio-CCR4-NOT complex promotes the translational repression of this mRNA, thus preserving germline self-renewal (Joly et al., 2013; Neumüller et al., 2008).

Also in *C.elegans*, *nos-1* and *nos-2* genes have been implicated in primordial germ cells (PGCs) development (Subramaniam and Seydoux, 1999). In particular, loss of *nos-1* and *nos-2* prevents germ cell proliferation during larval development. Similar to *Drosophila nanos*, *nos-2* is essential for PGCs colonisation of gonads, and *nos-1* is required for the viability of germ cells.

In *Xenopus,* one of the *nanos* genes, called *Xcat2*, was also shown to be involved in PGCs development. Indeed, *Xcat2* produces a protein which accumulates in the germplasm during PGCs formation (MacArthur et al., 1999). Interestingly, also in *Xenopus*, *Xcat2* is thought to promote the mitotic quiescence of germ cells, as in *Drosophila*. *Xcat2* (also called *Nanos1*) associates with the *cyclin* B mRNA *in vivo* and is thought to promote its translational repression in *Xenopus* germline (Lai et al., 2011).

In zebrafish, *nanos1* has also been shown to localise in the germplasm, and to be critical for the migration and viability of PGCs (Köprunner et al., 2001). Moreover, *nanos2* specifically labels germline stem cells (GSCs) in medaka fish ovaries (Nakamura et al., 2010), and is involved in ovary regeneration in zebrafish (Cao et al., 2019).

Therefore, *Nanos* genes have conserved functions during germ cell development, and interestingly, they often act as translational repressors of mRNAs, in order to promote mitotic quiescence (e.g., *cyclin* B) or suppression of apoptosis (e.g., *hid*).

Another critical function of *nanos* is the formation of embryonic polarity in *Drosophila* (Irish et al., 1989) (Figure 1.9). Although this role is not conserved in other organisms, it represents an established example of PTGR mediated by RBPs, and greatly contributed to the understanding of Nanos molecular mechanism.

Notably, *nanos* mRNA is a maternal factor deposited during oogenesis in the posterior part of the egg. Different RBPs, among which Rump and Aub, bind to specific sequences present in *nanos* 3' UTR (Bergsten and Gavis, 1999; Gavis and Lehmann, 1992, 1994), and allow *nanos* transcript localisation in the posterior pole of the egg (Becalska et al., 2011; Jain and Gavis, 2008). After fertilisation, *nanos* mRNA is translated and becomes expressed as a gradient, which increases from the anterior to the posterior pole. Then, Nanos, together with other two RBPs, Pumilio and Brat, recognise NREs in the 3' UTR of *hunchback* mRNA and repress it exclusively in the posterior pole (Hülskamp et al., 1989; Wharton and Struhl, 1991). In particular, Pumilio and Brat are the first to bind to the NREs of *hunchback*, and subsequently, Nanos joins the complex (Loedige et al., 2014; Sonoda and Wharton, 1999). Then, Nanos recruits CCR4-NOT, which promotes the deadenylation of *hunchback* mRNA (Wreden et al., 1997). Additionally, Brat recruits d4EHP to the 5’ cap of *hunchback*, and prevents the binding of eIF4E, thus inhibiting its translation (Cho et al., 2006).

In summary, an intricate network of RBPs and mRNAs, are responsible for the correct development of the abdomen, in the nascent *Drosophila* embryo.

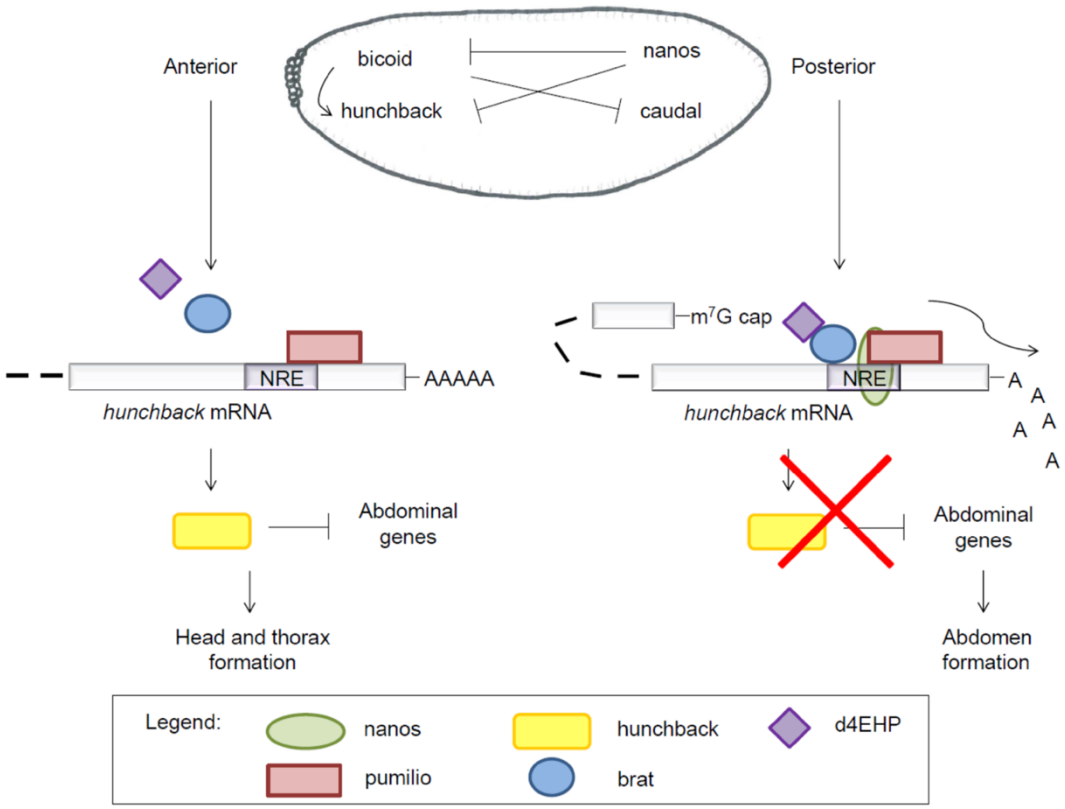


Figure 1.9. Example of PTGR mediated by Nanos in *Drosophila* embryos, adapted from (De Keuckelaere et al., 2018).

In the posterior pole of *Drosophila* embryos, Nanos binds to the RBP Pumilio and recognise *hunchback* mRNA, through the nanos response elements (NRE) present in the 3' UTR. Nanos and Pumilio, together with the RBP Brat, mediate the translational repression of *hunchback*, and determine the abdomen formation in the developing embryo.

1.8.3 The structure of NANOS proteins and its functional implications

As previously described, Nanos proteins have two conserved zinc fingers, (CCHC)2, within their C-terminal region. The first physical and chemical characterisations of this domain were performed by NMR and absorption spectroscopy, which determined that each CCHC motif in *Drosophila* Nanos bound to one zinc ion (Curtis et al., 1997). In particular, these data showed that Nanos is composed by two consecutive zinc finger motifs, C-X_2-C-X_{12}-H-

X_{10}-C and C-X_2-C-X_7-H-X_4-C, linked by a spacer, whose length (7 amino acids) is also highly conserved among Nanos orthologs. Moreover, the spacing between the zinc ligands in each motif seemed different from any other known zinc finger. Thus, Nanos domain seems to be a novel one, even distinct from the most similar retroviral CCHC domain.

Importantly, Curtis and colleagues also proved for the first time that the two zinc fingers were essential for Nanos function *in vivo* (Curtis et al., 1997). In particular, they showed that the cysteine residues (C) inside the zinc fingers were critical to rescue defects in *nanos* mutant embryos.

Moreover, they showed that Nanos was able to bind to RNA with high affinity *in vitro*, and this ability was dependent on the zinc fingers. However, the RNA-binding ability was not sequence-specific. Indeed, filter binding assays showed that mutations in the NRE sequence of *hunchback* mRNA did not significantly decrease Nanos binding.

Next, by performing a genetic screening aimed at identifying new *nanos* mutants, other authors showed that the zinc fingers were indeed necessary to support all *nanos* functions (Arrizabalaga and Lehmann, 1999). Moreover, in the same study, they discovered that another region, present after the zinc fingers (tail region) was also critical for at least two functions in *Drosophila*: migration of germ cells and embryonic polarity.

As previously mentioned, Nanos has a particular spacing between the cysteine (C) and histidine (H) residues inside each zinc finger. This feature makes Nanos zinc finger domain unique and led Hashimoto and colleagues to solve the crystal structure of the C-terminal region of zebrafish Nanos (Hashimoto et al., 2010a) (Figure 1.10). This revealed that Nanos zinc finger domain is composed of two independent lobes (motifs), which form a large cleft, and which are linked by a helix (constituted by the 7 amino acids spacer). In each lobe, three C and one H residues bind to one zinc ion, by tetrahedral

coordination. Importantly, this structure did not overlap with any other homologous domains, confirming that Nanos domain adopts a new structure (Hashimoto et al., 2010a). However, more recently, other structural studies revealed that the second zinc finger motif present in dNanos is homologous to the zinc knuckle structure of the HIV nucleocapsid protein (Weidmann et al., 2016) (Figure 1.10 E). Thus, Nanos zinc fingers might not be as unique as previously thought.

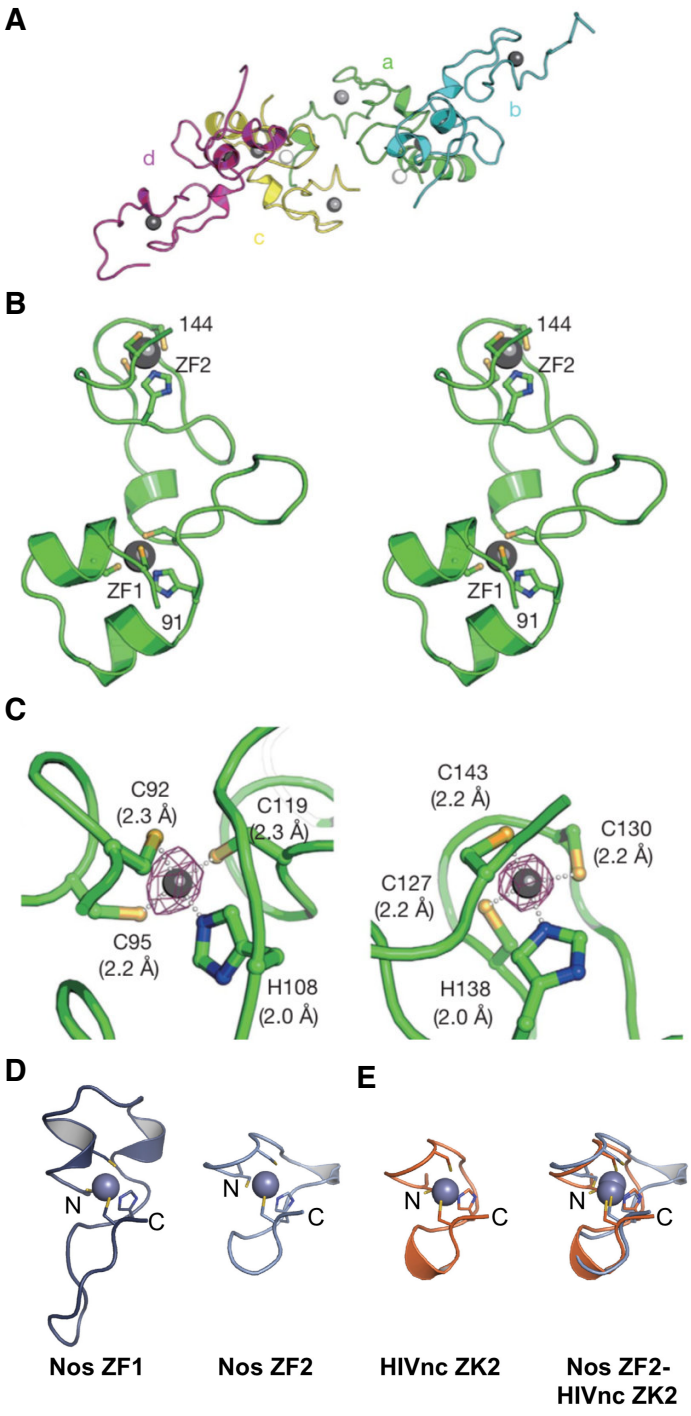


Figure 1.10. Structure of the zebrafish Nanos zinc-fingers and Zn knuckle (ZK) structure of the HIV nucleocapsid protein, adapted from (Hashimoto et al., 2010; Weidmann et al., 2016).

A. Ribbon diagrams of the zinc-fingers (ZFs) present in zebrafish Nanos (zNanos). Four molecules of zNanos are shown; zinc ions are grey and white spheres. **B.** Core part of the ZF1 and ZF2 are shown. Amino acid residues bound to the zinc ions (grey

spheres) are depicted by stick representation. **C.** The interaction between key amino acid residues (CCHC) of the ZF and the zinc ion is shown as white dots. Bond lengths are expressed in Ångström (Å). On the left it is shown ZF1, on the right ZF2. C is Cysteine, H is Histidine. **D.** Ribbon diagrams of the two ZFs. N and C are N- and C-terminal respectively. **E.** On the left, the Zn knuckle (ZK) structure of the HIV nucleocapsid protein (HIVnc) is shown. On the right, superimposition of the Nanos ZF2 and the HIVnc ZK.

Interestingly, several conserved amino acid residues in zebrafish Nanos crystal were found to be part of a large basic area, present on the surface of the two lobes (Hashimoto et al., 2010a). By mutating two aromatic residues included in this surface, Nanos zinc finger domain was shown to lose its RNA binding ability substantially, by EMSA. Moreover, the same mutations were previously found to cause strong *nanos* phenotype in *Drosophila*, suggesting an important link between *nanos* function and its RNA-binding ability.

Overall, these results also indicated, for the first time, that Nanos-RNA interaction was also mediated by electrostatic interactions between conserved basic amino acids and the phosphate blackbone of RNA.

Moreover, these authors confirmed by EMSA that Nanos zinc finger domain binds to RNA independently of the NRE sequences, as previously shown in 1997 by Curtis and colleagues. Thus, *Drosophila* Nanos binds to RNA with no sequence specificity.

As previously explained, Nanos binds to RNA together with Pumilio in *Drosophila*. To gain insight into the molecular basis of Nanos-Pumilio interaction with RNA, a new crystal structure was determined for the complex dNanos-Pumilio-RNA (Weidmann et al., 2016). In this crystal, dNanos zinc fingers were shown to enclose RNA, and both the zinc fingers and the C-terminal region of dNanos enveloped Pumilio.

Moreover, the addition of dNanos to the Pumilio-RNA complexes, induced conformational changes into Pumilio C-terminus, which resulted in the establishment of new interactions between Pumilio and RNA. In particular, in the presence of dNanos, Pumilio does not recognise a perfect Pumilio Response Element (PRE) anymore, but also contacts few nucleotides upstream. This was also confirmed by *in vitro* selection assays, which showed that dNanos could change the RNA motif bound by Pumilio, by enriching for A/U nucleotides, upstream of the PRE (Figure 1.11) (Weidmann et al., 2016).

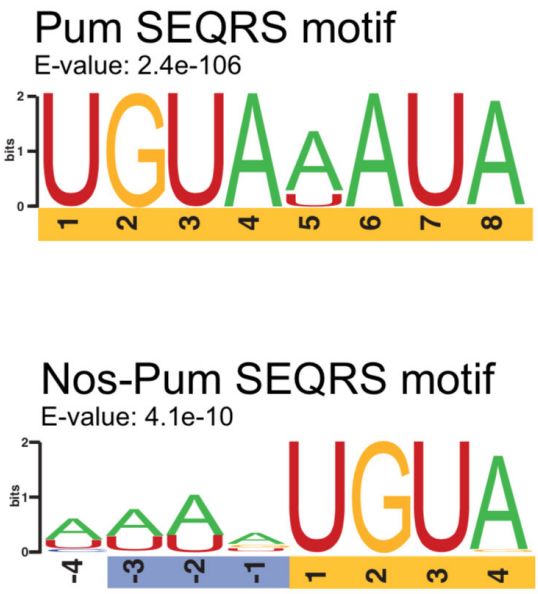


Figure 1.11. RNA-binding motifs of Pumilio (Pum) and Nanos-Pumilio (Nos-Pum) in *Drosophila*, adapted from (Weidmann et al., 2016).
The figure shows results from SEQRS (*in vitro* selection, high-throughput sequencing of RNA, and sequence specificity landscapes) analysis performed with Pumilio (top) and Nanos-Pumilio (bottom). When Nanos protein is added to Pumilio, A/U-rich sequences are enriched upstream of the Pumilio response element.

Weidmann and colleagues also showed for the first time that dNanos mediates sequence-specific interactions with RNA, at the atomic level. In particular, this

interaction occurs between three amino acids present in a hydrophobic binding pocket in dNanos and the third nucleotide of the 'AAU' Nanos binding site (NBS) (which is upstream of the PRE). Importantly, the same three amino acids are also essential to repress a luciferase reporter containing the NRE sequence (Weidmann et al., 2016), thus highlighting the link between Nanos repressive function and its RNA-binding ability.

Additionally, when dNanos RNA-binding ability was assessed by EMSA, dNanos could shift RNA only when in complex with Pumilio, but not alone (Weidmann et al., 2016). This result suggested that dNanos does not bind to RNA independently, but only together with other RBPs.

Besides the zinc finger, Nanos proteins also have another conserved region, which is used to interact with the CCR4-NOT (CNOT) complex, in order to recruit it to various RNAs (Suzuki et al., 2012; Wreden et al., 1997). In vertebrates, Nanos proteins interact with CNOT through the Nanos Interacting Motif, NIM, which is only 17 aa long, and is located within a disordered region of Nanos N-terminus (Bhandari et al., 2014). The NIM directly interacts with the C-terminal SHD domain in CNOT1. Importantly, when key aromatic residues are mutated inside the NIM, Nanos cannot repress RNA anymore *in vitro*.

In contrast, invertebrates do not have a NIM sequence. For instance, *Drosophila* Nanos has the Nanos effector domain (NED), which is responsible for the interaction with CNOT (Raisch et al., 2016). The NED is significantly longer than the NIM (187 aa) and seems to contact the C-terminal region of CNOT1-3 multiple times. Importantly, also in *Drosophila*, this domain was shown to be essential for Nanos repressive activity on RNA.

In summary, these structural studies illustrated the molecular basis of the interaction between Nanos and RNA, and also between Nanos and its protein

partners. Moreover, they clarified the molecular basis of various Nanos phenotypes.

1.8.4 *Nanos2* in male gonocytes

The first *Nanos* gene identified and studied in mouse was called *Nanos1*. Similar to its orthologs, *Nanos1* was shown to be present in mouse oocytes and to be maternally supplied (Haraguchi et al., 2003). However, in contrast with the restricted expression pattern of other *Nanos* orthologs, *Nanos1* is evenly localised in the cytoplasm of mouse oocytes and, after fertilisation, the maternal copy is rapidly replaced with the zygotic one. Moreover, *Nanos1* is not expressed in PGCs, but in adult mice, it is present during oocyte maturation and in spermatids. However, *Nanos1*$^{-/-}$ mice are completely fertile, indicating that, in contrast with its orthologs, *Nanos1* is dispensable for germ cell development. Interestingly, *Nanos1* expression was also detected in the central nervous system, although its loss does not cause morphological or behavioral defects (Haraguchi et al., 2003).

Next, other two *Nanos* paralogs were identified in mouse and called *Nanos2* and *Nanos3.* These are two zygotic genes which were found to be essential for germ cell development (Tsuda et al., 2003). Thus, although PGCs specification significantly differs among different organisms (with *Drosophila*, *Xenopus,* and zebrafish relying on maternal determinants, whereas the mouse does not), *Nanos* family has a conserved evolutionary function in germ cells.

Both *Nanos2* and *Nanos3* are present in mouse gonocytes, although they have different expression profiles and thus distinct phenotypes (Tsuda et al., 2003). *Nanos2* mRNA is restricted to male gonocytes, from E13.5 to E.16.5 (see Figure 1.1 A), and the genetic deletion of both copies of *Nanos2* causes exclusively male infertility.

In contrast, *Nanos3* mRNA is expressed earlier than *Nanos2* in PGCs, from E9.5, when PGCs have not undergone sex determination (Figure 1.1 A). Also,

Nanos3 is detected until E14.5 in male gonocytes, but only until E13.5 in females, and the genetic deletion of *Nanos3* causes both male and female infertility.

The phenotype of *Nanos2*$^{-/-}$ mice is already detectable at E15.5, as some germ cells are mislocalised outside of the tubules at this stage. Additionally, at E18.5, the number of germ cells is significantly reduced in *Nanos2*$^{-/-}$ testes, and these become devoid of germ cells in 4-weeks old mice. Indeed, these mice showed a decrease in testicular weight by a factor of 70%. Importantly, from E15.5 onwards *Nanos2*$^{-/-}$ germ cells were found to be apoptotic, suggesting that *Nanos2* in mouse has a conserved function in the suppression of apoptosis, like in *Drosophila*.

In contrast, in *Nanos3*$^{-/-}$ mice, both ovaries and testis do not contain germ cells already at E15.5. This is due to a defect in PGCs maintenance during their migration towards the gonads, rather than to a defect in PGC specification (Tsuda et al., 2003).

Although the loss of *Nanos2* or *Nanos3* causes different phenotypes, it was also shown that NANOS2 and NANOS3 proteins cross-talk with each other, and they also have partially redundant functions (Suzuki et al., 2007). Indeed, NANOS2 and NANOS3 expression overlap in a narrow window of time. NANOS3 protein is mainly expressed between E12.5-E13.5, but greatly decreases from E14.5 and is almost non-detectable in E15.5. However, NANOS2 is expressed from E13.5, and its level remains constant until E17.5. Thus, NANOS2 overlaps with NANOS3 between E13.5 and E14.5. Interestingly, in *Nanos2*$^{-/-}$ germ cells, NANOS3 expression is upregulated from E13.5, both at the protein and mRNA level. This suggests a feedback mechanism, where, in the absence of NANOS2, some putative positive regulators of NANOS3 would be expressed at higher levels, thus enabling NANOS3 expression (Suzuki et al., 2007).

Interestingly, another type of feedback mechanism was also observed for NANOS2 itself, which seems to control its own expression level. Indeed, in

Nanos2-overexpressing mice, NANOS2 endogenous level decreases, probably due to a regulation occurring at the post-transcriptional level, through Nanos2 3' UTR (Suzuki et al., 2007).

The redundancy of *Nanos2* and *Nanos3* genes was investigated by performing rescue experiments in mouse (Suzuki et al., 2007). Importantly, *Nanos2* was shown to rescue defects in *Nanos3*$^{-/-}$ mice, although *Nanos3* could not substitute for *Nanos2* (Suzuki et al., 2007).

Indeed, following *Nanos3* expression in *Nanos2*$^{-/-}$ mice, the embryonic gonads still display mislocalised germ cells and many apoptotic cells. Moreover, 4-weeks old mice have no germ cells left in testes. Overall, these findings highlighted that *Nanos2* also has additional and specific functions independent of *Nanos3*, throughout male germ cell development.

The precise functions of *Nanos2* were further investigated also by other groups. For instance, it was shown that *Nanos2* is required to maintain the mitotic quiescence of male germ cells (Suzuki and Saga, 2008), although it is not essential for quiescence initiation (Saba et al., 2014). Indeed, male germ cells typically enter G0/G1 quiescence at E14.5, but, in *Nanos2*$^{-/-}$ mice, male gonocytes abnormally resume proliferation from E15.5.

Moreover, NANOS2 was shown to have other two functions. First of all, *Nanos2* is essential to prevent male gonocytes from entering meiosis (Suzuki and Saga, 2008). It was shown that *Nanos2*$^{-/-}$ germ cells up-regulate Stra8 from E14.5, which might be the cause of premature meiosis induction.

Moreover, Nanos2 is also required to induce the male gene expression program in male germ cells, which otherwise retain an undifferentiated gene expression profile, similar to PGCs (Saba et al., 2014). Importantly, this function was shown to be independent of meiosis suppression.

Furthermore, scientists sought to investigate the molecular mechanism underlying *Nanos2* functions. Notably, it was proposed that NANOS2 suppresses meiosis through the post-transcriptional repression of few

transcripts (Suzuki et al., 2010). This is thought to be mediated by direct recruitment of CNOT to NANOS2 mRNAs targets, which subsequently would cause RNA degradation. This model is supported by experimental data, showing that NANOS2 physically interacts with CNOT1 in male gonads and colocalises with the CNOT complex in processing bodies (PBs) (Suzuki et al., 2010). Additionally, native RIP experiments revealed that NANOS2 interacts with some mRNAs involved in meiosis, such as *Sycp3*, *Stra8*, *Taf7l*, *Dazl*, and Meisetz, which are also up-regulated in *Nanos2*$^{-/-}$ gonads. Importantly, the direct interaction between the first ten amino acids of NANOS2 with the C-terminal of NOT1 is essential for NANOS2 function in male gonocytes (Suzuki et al., 2012a).

Moreover, NANOS2 was suggested to inhibit the meiotic entry of male germ cells by antagonising DAZL expression in male gonocytes (Kato et al., 2016). Notably, NANOS2 was shown to bind to *Dazl* 3' UTR and to stimulate its repression.

Furthermore, similarly to *Drosophila* NANOS, which works together with the RBP PUMILIO, NANOS2 was shown to cooperate with DND1, another RBP, in mouse gonocytes (Suzuki et al., 2016). In particular, NANOS2 interacts directly with DND1 in male gonocytes, through its zinc finger domain, which surprisingly, was shown to mediate protein-protein interactions and not only RNA-binding. Thus, in the male gonad, NANOS2 is thought to achieve targets' selection, by binding to only a subset of mRNAs, such as *Sycp3*, *Dazl*, and *Nanog*, in presence of DND1. This interaction would then trigger their degradation through the action of the CNOT complex (Suzuki et al., 2016).

1.8.5 *Nanos2* in adult SSCs

As previously explained, *Nanos2* is essential for male germ cells development in the embryonic gonad (Tsuda et al., 2003). However, *Nanos2* has also a

separate function in the maintenance of adult SSCs (Sada et al., 2009). NANOS2 was shown to be expressed in the undifferentiated spermatogonia compartment, mainly in the As and Apr cells, marked by the self-renewing marker GFRα1 (Sada et al., 2009; Suzuki et al., 2009). Importantly, conditional deletion of the *Nanos2* gene in post-natal SSCs revealed that germ cells were gradually lost in the adult testis (Figure 1.12). Notably, 8-weeks-old mice displayed a great reduction of germ cells in the testes, and 12-weeks-old mice had completely lost germ cells (Sada et al., 2009).

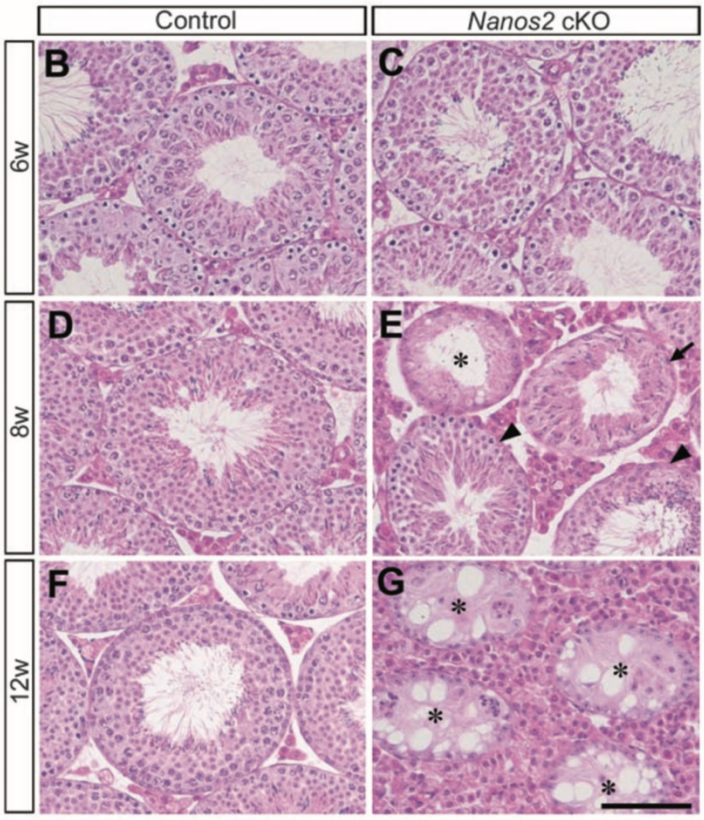


Figure 1.12. Histological analysis of post-natal mouse testes lacking *Nanos2*, adapted from (Sada et al., 2009).

6-, 8- and 12-weeks-old testes of Control and *Nanos2* cKO (conditional knockout) mice were analysed by immunostaining of testicular cross-sections with Hematoxylin and Eosin. In 8-weeks-old-*Nanos2* cKO testes (**E.**) many tubules do not have spermatogonia (arrowheads), and some of them have only elongated spermatids (arrow). By 12 weeks (**G.**), tubules lack all germ cells (asterisks).

In particular, following *Nanos2* deletion, GFRα1 SSCs were depleted, indicating that the SSCs pool was the main one affected among the spermatogonia populations and that it caused germ cell loss over time.
Additionally, lineage-tracing experiments revealed that *Nanos2*-expressing cells can give rise to long-term patches containing all the spermatogenic lineages (Sada et al., 2009). Importantly, the number of these patches are ten times higher than those formed by *Ngn3*-expressing cells, which are a transit-amplifying population (Nakagawa et al., 2007, 2010; Yoshida et al., 2007). Thus, *Nanos2*-expressing cells are able to self-renew in the long term as well as to differentiate, in order to produce all the different stages of spermatogenesis. Thus, *Nanos2*-expressing cells might be more enriched for the actual stem cell population, (as GFRα1). Moreover, when *Nanos2* is overexpressed in mouse, undifferentiated spermatogonia greatly increase in number, at the expense of the differentiating pool. This accumulation is due to a block in SSCs differentiation and not to increased proliferation rates of SSCs.

Despite this extensive characterization of *Nanos2* function in adult testes, the mechanism through which NANOS2 maintains SSCs has not been widely studied. However, some insight comes from Zhou and colleagues (Zhou et al., 2015). NANOS2 was shown to localise to PBs in male gonocytes (Suzuki et al., 2010) and similarly, it also co-localises with Rck, a marker for PBs and stress granules, in post-natal SSCs *in vivo* (Zhou et al., 2015). Additionally, by promoting mRNPs assembly in SSC lines, NANOS2 is thought to indirectly repress various differentiation genes, such as *Sohlh1*, *-2*, *Ngn3*, and *c-Kit*,

which would be trapped and stored in these granules, thus allowing SSCs self-renewal.

It was also proposed that NANOS2 directly represses other mRNAs involved in SSCs differentiation, such as *Sohlh2*, *Dazl*, and *Dmrt1*, also by stimulating their degradation (Zhou et al., 2015). Although his mechanism was never fully demonstrated, other data previously showed that NANOS2 has deadenylase activity *in vitro*, when it is co-immunoprecipitated with CNOT, from NANOS2-overexpressing adult testis (Suzuki et al., 2010).

Moreover, NANOS2 was shown to sequester mTOR (Zhou et al., 2015), the catalytic subunit of the mTORC1 complex, in mRNPs in SSC lines. mTORC1 is known to be activated by retinoic acid (Lal et al., 2005) and to promote SSCs differentiation (Hobbs et al., 2010). Thus, the authors suggested that NANOS2 represses SSCs differentiation also by relegating mTOR to mRNPs. Moreover, mTORC1 signaling stimulates cell growth and protein translation (Laplante and Sabatini, 2012). Since NANOS2 targets are thought to be repressed post-transcriptionally, NANOS2 would ensure the translational repression of differentiation genes, also by repressing the mTORC1 pathway (Zhou et al., 2015).

Furthermore, it was recently shown that the transition from self-renewing to differentiating SSCs is controlled by modulating NANOS2 expression at the protein level (Zhou et al., 2017). Upon exposure to differentiating signals, such as retinoic acid, SSCs induce NDFIP2 expression, an activator of the E3 ubiquitin ligase, NEDD4. NDFIP2 stimulates the NEDD4-dependent ubiquitylation of NANOS2, which results in NANOS2 degradation and allows SSCs differentiation. Interestingly, NEDD4 was also shown to protect SSCs under heat-stress conditions, by controlling mRNPs clearance and an excessive accumulation of misfolded proteins and stress granules in SSCs. Therefore, a single protein, NEDD4, can promote SSCs differentiation by removing NANOS2 under homeostatic conditions, and at the same time enhances SSCs survival under heat-stress conditions.

1.9 Methods to detect RNA-protein interactions

In the past decade, hundreds of new RBPs have been identified by genome-wide approaches in yeast, mouse, and human cell lines (Baltz et al., 2012; Beckmann and Granneman, 2019; Beckmann et al., 2015; Castello et al., 2012; He et al., 2016; Shchepachev et al., 2019; Urdaneta et al., 2019). These studies revealed that many of the new RBPs were well-studied metabolic enzymes (Beckmann et al., 2015), or chromatin-associated factors (He et al., 2016). Overall, this suggests that the biological importance of RBPs might have been underestimated for a long time. Therefore, a detailed characterisation of both known and newly identified RBPs might reveal unprecedented connections among different and unrelated biological processes.

The first step towards the understanding of RBPs' function necessarily implicates the investigation of their RNA-protein interactions. Thus, approximately thirty biochemical methods have been developed during the past twenty years to detect RBP-RNA interactions, and these are reviewed in (Lee and Ule, 2018). Additionally, to gather the knowledge accumulated on RBPs-RNA interactions, thanks to the application of these techniques, a new database was created, and this was called 'ATtRACT' (Giudice et al., 2016).

One of the first attempts to characterise RNA-protein interactions was aimed at identifying the RNAs bound to the spliceosomal proteins, from patients' samples with the lupus erythematosus disease (Lerner and Steitz, 1979). To this end, the protein of interest was pulled down with antibodies, and the bound RNAs were characterised. The particular abundance of the splicing protein studied allowed the identification of its targets with such a non-sensitive method. This type of approach, known as RIP (RNA and Immunoprecipitation), is performed under native conditions, which means that RNA-protein complexes are not stabilised by any strong covalent bond, before the cell lysis. However, there are major concerns regarding the specificity of the RNAs

retrieved in this way. Indeed, non-physiological interactions can also occur during and after the lysis, when all the different cellular compartments are mixed, and RBPs can potentially re-assemble in new complexes, with non-physiological targets (Mili and Steitz, 2004). Thus, other methods have been developed to stabilise RBPs-RNA complexes before the cell lysis, by taking advantage of the cross-linking (Niranjanakumari et al., 2002). However, the use of formaldehyde to induce the cross-linking might also lead to false-positive RNA-protein interactions. Indeed, formaldehyde not only cross-links nucleic acids to proteins, but it also induces covalent bonds between different proteins. This results in the purification of RNA potentially associated with multiple RBPs, leading to possible misleading conclusions regarding the interaction between a single RBP and its targets.

In order to study direct protein-RNA interactions, scientists developed the UV cross-linking and immunoprecipitation, CLIP, where they employed the UV-C light as a cross-linking agent (Ule et al., 2003, 2005; Urlaub et al., 2000). The advantage of the UV-C light (254 nm) is that it mainly cross-links RNA to proteins and that it only cross-links amino acids and nucleotides which are at 'zero-distance' (order of Ångströms apart) (Greenberg, 1979). Importantly, this greatly increases the sensitivity of the method, as also transient, but physiological, RNA-protein interactions are captured, and not only the most abundant RNA species are precipitated. Additionally, the presence of irreversible covalent bonds between RNA and proteins enables their purification under semi-denaturing conditions. This significantly decreases the RNA background and the possibility of co-purifying other RBPs, thus increasing the specificity of the technique.

After immunopurifiying the RBP of interest, RT-qPCR was initially employed to determine the identity of the RNAs cross-linked to the protein. However, this impaired the characterisation of RNA occupancy at the transcriptomic level. Thus, scientists started to couple RIPs with microarrays and then with sequencing, (RIP-Chip and RIP-seq) (Keene et al., 2006; Tenenbaum et al.,

2000; Zhao et al., 2010), whereas CLIP was coupled directly to high-throughput sequencing, (HITS-CLIP) (Licatalosi et al., 2008).

Although CLIP greatly improved the characterisation of RBP-RNA interactions, it also has some drawbacks. For instance, RNA-proteins complexes can be purified only under semi-denaturing conditions and it relies on the existence of specific antibodies towards the protein of interest. Even when these are available, they can still lead to the isolation of many false-positive RNAs. In order to limit these problems, an alternative approach was developed, the UV-Cross-linking and Analysis of cDNA (CRAC) (Granneman et al., 2009a). Here, the RBP of interest is modified in order to be fused to a complex tag, which includes at least three parts: an epitope tag, a cleavage site, and a 6xHis sequence. First, the epitope tag is used for the immunopurification of the protein with an antibody. Next, a cleavage site allows the enzymatic removal of the previous tag and the bound antibody, after immunoprecipitation. Then, a 6xHis affinity tag allows stringent purification of the RNA-protein complexes on Nickel columns, under denaturing conditions. Overall, this method greatly reduces the number of false-positive RNA-protein interactions. However, the addition of such a tag can potentially affect the biology of the RBP. Thus, appropriate analyses need to be done before performing CRAC, to ensure that the fusion protein is fully functional.

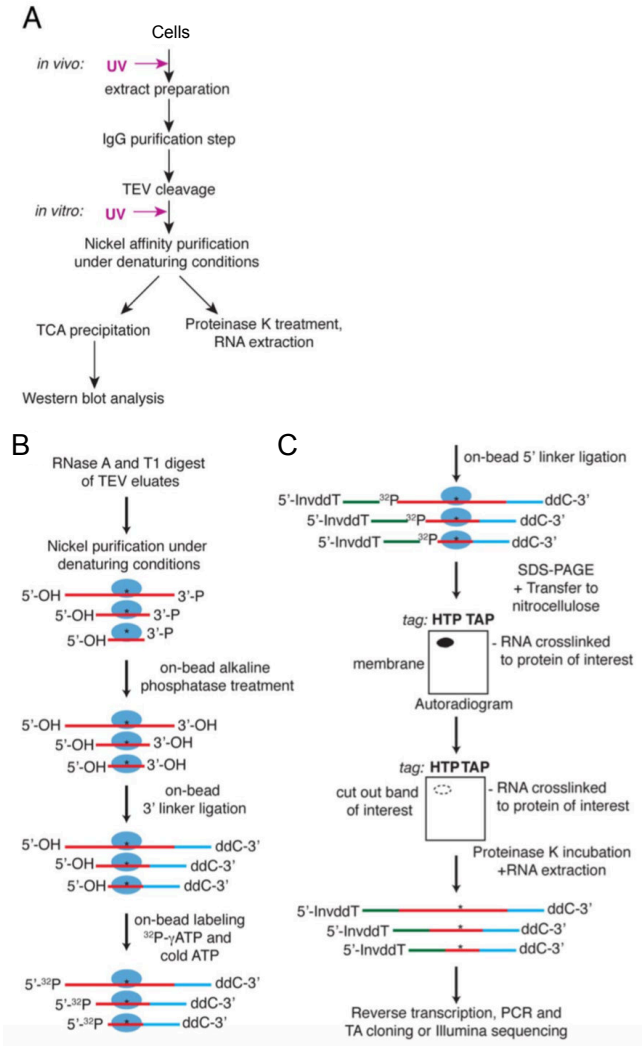


Figure 1.13. Schematic representation of the CRAC protocol, adapted from (Granneman et al., 2009a).

A. Recombinant proteins are used for CRAC. The tag is fused to the N- or C-terminal of the protein of interest and includes an epitope for immunopurification, a PreScission site for protease cleavage and a 6xHis tag for affinity purification on nickel beads. Proteins are permanently cross-linked to RNA by 254 nm UV-light and cells are lysed. RNA-protein complexes are immunopurified by using an antibody against the tagged-protein. The antibody is removed through protease cleavage (TEV), which cleaves the PreScission-tagged protein. RNA-protein complexes are then purified through nickel affinity purification thanks to the 6xHis tag present in the recombinant protein. Denaturing washes are applied to reduce the RNA background. After chemical modification of the cross-linked RNA (see **B.**), proteins are digested by Proteinase K and RNA is extracted for further analysis. For quality control, after eluting proteins

from nickel columns, a fraction of the eluate can also be used for protein precipitation and analysis by western-blot (WB). **B.** After protease cleavage, cross-linked RNA is partially digested by RNase A and T1. Denaturing washes are applied, while RNA-protein complexes are still immobilised on nickel beads. RNAs is treated with phosphates, to remove the phosphates left by RNase. A 3' adapter (shown in light blue) is ligated to RNA, which is also radio-labelled with ^{32}P-γ-ATP. **C.** A 5' adapter (in green) is ligated to RNA and RNA-protein complexes are eluted from beads by using imidazole. After resolving RNA-protein complexes by using SDS-PAGE, they are transferred to a membrane, which is analysed by autoradiography. Areas corresponding to the mass of the radio-labelled protein are excised from the membrane and subjected to Proteinase K treatment. RNA is extracted and reverse transcribed. Readthrough cDNAs with deletions (asterisks), which correspond to cross-linking sites, are amplified by PCR and libraries are sequenced.

In Figure 1.13, detailed steps of the CRAC protocol are shown. Proteins are permanently cross-linked to RNA by using UV-C light on live cells. After cell lysis, RNA-protein complexes are immunoprecipitated (IP) with an antibody against a tag (Figure 1.13 A). To further purify the cross-linked RNA-protein complexes, the tag used for the IP is enzymatically removed by using a protease (TEV or 3C). This enzyme specifically cleaves between glutamine and glycine residues present in the PreScission sequence of the PreScission-tagged RBP. By removing the IP tag, also the bound antibody is displaced, which is important in order to reduce non-specific protein and RNA contaminants which might associate with IgG.

Afterwards, the cross-linked RNAs are partially digested with RNase A and T1. RNA digestion decreases the chances of co-purifying additional RBPs, which might reside on the same RNA molecule, and also provides RNA with a size compatible for sequencing.

To allow very stringent purification of the RNA-protein complexes, 6xHis-tagged RBPs are immobilised on nickel beads, thanks to the strong affinity between nickel and His residues. Non-specifically bound RNAs are not cross-

linked to the RBPs and thus are washed away from the immobilised RBP by applying denaturing buffers. This is a critical step which greatly decreases the RNA background.

While being on nickel beads, RNase-digested RNAs are chemically modified to allow incorporation of the 3' adapter, radiolabelling, and ligation of the 5' adapter. First, RNAs are 3' dephosphorylated to remove the phosphate group previously left by RNase. Then, RNAs are ligated to the 3' adenylated adapter, which is complementary to the reverse transcription primer used later on. Afterwards, RNAs are radio-labelled with ^{32}P-γ-ATP and a 5' adapter is ligated to the RNAs. The 5' adapter includes an experimental barcode which has a variable sequence length (2-6 bases) and composition. Since each sample is labelled with a different experimental barcode, different libraries can be pooled and sequenced together. The 5' adapter contains also a 3-bases UMI (unique molecular identifiers), to allow identification of PCR duplicates by bioinformatic analysis, at a later stage. A schematic representation of the adapters and primers used in CRAC is shown in Figure 1.14.

After modifying the RNAs, RNA-protein complexes are eluted from nickel beads by using imidazole, which competes with the imidazole ring of His (in the 6xHis-tagged protein) for the binding to nickel beads.

Eluted RNA-protein complexes are resolved on a gel through SDS-PAGE separation. and transferred on a membrane. During this step free and non-specific RNAs are further removed, since they cannot migrate on the membrane. Visualisation of the radiolabelled-RBP is crucial to verify whether the RNA fragmentation was successful, and also the correct size of the purified protein-RNA complexes, which should migrate above the free protein. This step is critical also to exclude the presence of other bands at different sizes, which would correspond to other RBPs.

Then, RNA-protein complexes are excised from the membrane, and proteins are digested by Proteinase K to isolate RNA and perform reverse transcription (RT). Importantly, the amino acid residue that was originally cross-linked to the interacting RNA by UV light cannot be digested due to the covalent bond with

a ribonucleotide. Therefore, this residue remains attached to RNA and obstructs the transition of the enzyme during RT. If the enzyme overcomes the cross-linked residue on the RNA fragment, it generates a readthrough cDNA which often has mutations at these cross-linking sites.
After reverse transcription, a library is generated from the cDNA by PCR amplification and purified libraries are eventually sequenced.

Sequencing reads containing mutations can be used to map the original protein binding site on RNA. Importantly, substitutions and deletions seem to be the direct consequence of the reverse transcriptase acting at the cross-linking site, rather than sequencing or alignment errors (Granneman et al., 2009a). Moreover, the identification of the binding site in readthrough cDNAs is likely to be more accurate when deletions are used (Zhang and Darnell, 2011).

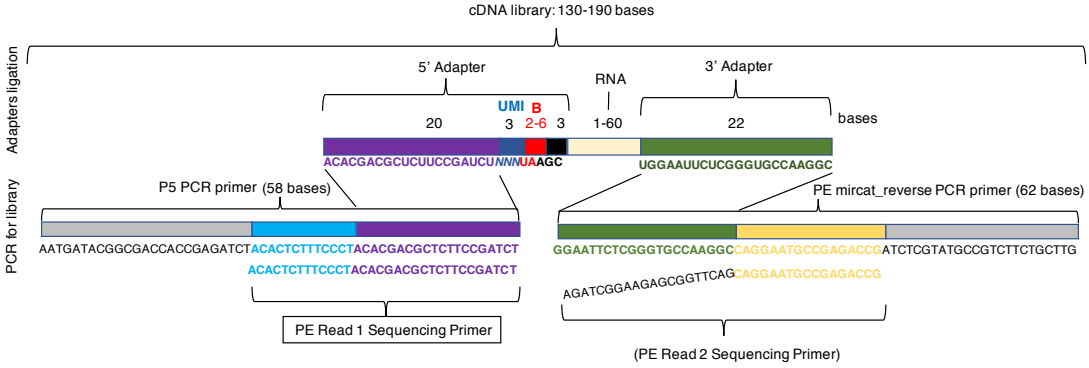


Figure 1.14. Schematic representation of the adapters and primers used to generate CRAC libraries.

CRAC libraries are ~ 130-190 bases in size. The RNA insert included between the adapters, ranges between 1 and 60 bases, depending on the efficiency of the RNase

treatment. 5' and 3' adapters are ligated before reverse transcription. The 5' adapter carries a particular sequence, the experimental barcode ('B', in red), which is 2-6 bases long. It also includes a 3-bases UMI (unique molecular identifiers). RNA is reverse transcribed by using a primer complementary to the 3' adapter. For library generation, cDNA is amplified by PCR by using primers which anneal to the 5' and 3' adapters. Since only the 5' 'PE Read1 Sequencing Primer' anneals perfectly to the library, CRAC libraries are sequenced from the 5' end (single end sequencing).

Like CRAC, CLIP allows detecting the protein binding site on RNA at single-nucleotide resolution, by using mutations (deletions) present in readthrough cDNAs (Lee and Ule, 2018; Ule et al., 2003).

Interestingly, a variant of CLIP called individual-nucleotide resolution UV cross-linking and immunoprecipitation (iCLIP) was developed to enable mapping of protein binding sites by using both truncated and readthrough cDNAs (König et al., 2010). Truncated cDNAs are produced whenever the reverse transcriptase stops at cross-linking sites (Figure 1.15, step 8). At these positions, the cross-linked amino acid residue constitutes an obstacle for the enzyme, which can frequently fall off the cDNA thus generating a truncated molecule. Truncated cDNAs are therefore indicative of the cross-linking sites. However, they are not retrieved in CLIP and CRAC protocols because they lack the SeqFw adapter at the 5' end, which was incorporated before reverse transcription (RT). Therefore, only full-length (readthrough) cDNAs have the adapter and can be further amplified and sequenced.

Differently from CRAC and CLIP, in iCLIP truncated cDNAs can be retrieved because the SeqFw adapter is incorporated at the 3' end, during RT through the RT primer. Therefore, despite truncated cDNAs are shorter than readthrough cDNAs, they still have both adapters at the 3' end and this allows their amplification later on. After RT, purified cDNAs are circularised through an enzymatic reaction, and circular cDNAs are linearised to generate libraries for sequencing. Further details of the iCLIP method are shown in Figure 1.15.

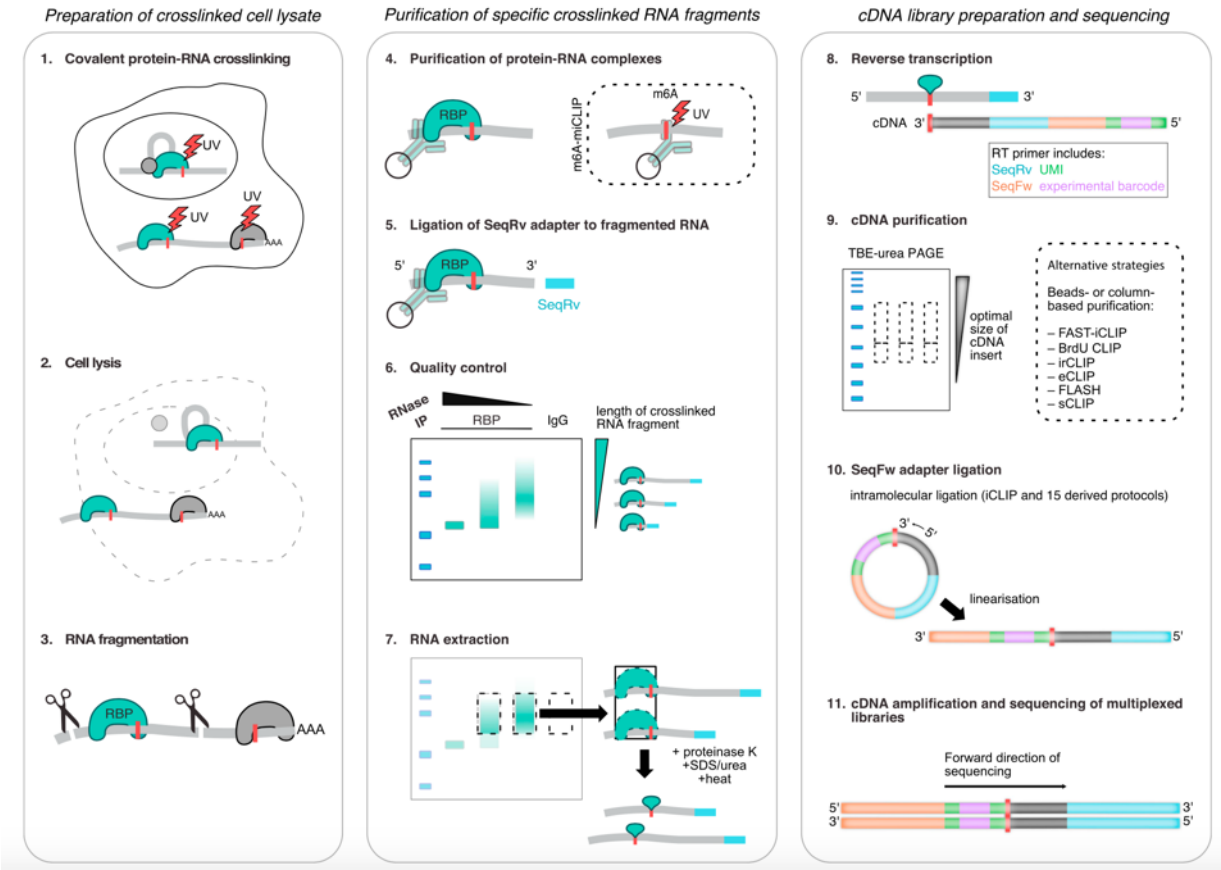


Figure 1.15. Schematic representation of the main steps performed in iCLIP (Lee and Ule, 2018).

The first part of iCLIP (1-4) is performed as in CLIP. However, other steps have been modified to map the protein binding site through the analysis of truncated cDNA molecules. In iCLIP, truncated cDNAs are retrieved because they contain the SeqFW adapter, which is incorporated during RT through the RT primer, at the 3' end (8). The RT primer also includes unique molecular identifiers (UMI) and an experimental barcode. After purification (9), the cDNA is circularised by using the circligase and then linearised thanks to a restriction site (10). cDNA is amplified by PCR to generate a library (11) and subsequently sequenced.

Alternative ways to cross-link proteins to RNA have also been employed and rely on the incorporation of nucleoside analogues (s^4U) into RNA. These analogues are then covalently linked to proteins, upon exposure of cells to the UV-A light (365 nm) (Hafner et al., 2010). Interestingly, this wavelength seems to capture better certain RBPs-RNA interactions than the UV-C light at 254 nm (Lee and Ule, 2018). However, these methods are partially limited by the incorporation rate of the nucleoside analogues in cells and tissues, and by the cytotoxicity of the same compounds.

As in CLIP and CRAC, also in PAR-CLIP readthrough cDNAs are used to detect the protein binding site on RNA (Hafner et al., 2010). Notably, in PAR-CLIP the identification of thymidine to cytidine mutations present in the readthrough cDNA can determine the initial s^4U positions, which represent the the cross-linking sites.

More recently, another category of methods has been developed to detect RNA-proteins interactions *in vivo*, and these are RNA tagging (Lapointe et al., 2015) and TRIBE (McMahon et al., 2016). The approach used is similar to DamID (Steensel and Henikoff, 2000), which allows determining the genomic occupancy of transcription factors *in vivo*. In RNA tagging and TRIBE, the RBP of interest is engineered in order to be fused to the catalytic domain of an RNA editing enzyme. In this way, the enzyme is recruited by the RBP to its RNA targets and edits them with permanent chemical modification. Thus, RNA editing substitutes for the UV-cross-linking, which is used in CLIP for the same purpose. This represents a great advantage since RNA editing occurs *in vivo* and significantly reduces the input material of the experiment. Thus, RNA-protein interactions can also be captured in rare populations of cells, such as 150 neurons (McMahon et al., 2016).

Notably, RNA tagging was developed in yeast, and it is based on the addition of the catalytic subunit of an oligo(U) enzyme to the RBP of interest (Lapointe et al., 2015). TRIBE was developed in *Drosophila*, and it combines the catalytic domain (cd) of another RNA-editing enzyme, cdADAR, to RBPs, in order to mark their RNA targets (McMahon et al., 2016). When TRIBE was compared to CLIP, it showed lower sensitivity, as it seems to detect by far fewer targets. This aspect was improved by incorporating a mutation in the cdADAR to enhance the RNA editing efficiency and at the same time to reduce the sequence bias of the enzyme for certain nucleotides (HyperTRIBE) (Rahman et al., 2018; Xu et al., 2018).

However, these *in vivo* methods have some limitations. First of all, they do not allow to detect the exact binding site of RBPs on RNA, whereas CLIP enables its detection at single-nucleotide resolution. Moreover, the function of the fusion protein needs to be assessed before performing the experiment, to exclude that the enzyme does not negatively impact the RBP.
Importantly, RNA editing enzymes might exert toxic effects in specific cell types or during certain developmental stages. Thus, to achieve the appropriate, non-toxic, expression levels, it might be necessary to use conditional genetics, and this can be complicated when applied to mammalian systems *in vivo*.
In summary, future refinements of these sophisticated '*in vivo* methods', might lead to their broad applications also to other organisms, besides yeast and *Drosophila*.

Aims of the thesis

In this thesis, I dissected the molecular function of the RNA-binding protein (RBP) NANOS2 in mouse SSCs. In order to understand the biology of any RBP, it is imperative to define its actual RNA interactome. Other scientists have previously sought to determine the molecular mechanism underlying NANOS2 function and, notably, they attempted to characterise its targets in SSC lines, by using native RNA immunoprecipitation (Zhou et al., 2015). However, this technique has some disadvantages, which do not allow identifying the RNA occupancy of an RBP unambiguously (Lee and Ule, 2018; Mili and Steitz, 2004). Thus, the first aim of this work was to characterise NANOS2 RNA targets transcriptome-wide, in SSC lines. To do so, we employed one of the most rigorous methods, called CRAC (Granneman et al., 2009b). Hence, we engineered a *Nanos2*^{TAG} mouse allele, which we used to derive SSC lines and to perform CRAC. Therefore, we addressed whether NANOS2-RNAs interaction is direct and which types of RNAs are bound by NANOS2. Moreover, we precisely defined NANOS2 binding site on RNA, at the single-nucleotide resolution. After identifying the biological processes where NANOS2 targets are involved in, we also sought to determine their expression profile throughout adult spermatogonia development.

Next, this thesis aimed to characterise NANOS2 molecular function in mouse SSCs by identifying its protein interactor. To address this, we performed immunoprecipitation followed by mass spectrometry in SSC lines, which revealed that many subunits of the CCR4-NOT complex were the most confident NANOS2 interactors.

The third objective of this work was to determine the molecular consequences of CNOT-NANOS2 interaction. Since CNOT is the major deadenylase in eukaryotic cells and it promotes RNA decay (Collart and Panasenko, 2012; Shirai et al., 2014), my work aimed to investigate the stability of NANOS2 mRNA targets in SSC lines, by using a novel technique called SLAM-seq (Herzog et al., 2017).

Overall, these findings significantly expanded our knowledge on how the evolutionarily conserved RBP NANOS2 works at the molecular level and sheds light on how NANOS2 can maintain SSCs self-renewal in the mouse.

Chapter 2. Materials and methods

2.1 Mouse generation and maintenance

2.1.1 Mouse strains/backgrounds and diet

All mice were maintained on the C57BL/6 genetic background. Mice from *Nanos2*^{TAG} line were also bred into DBA/2J genetic background (purchased from the Charles River Laboratories) to generate F1 hybrid mice with a mixed C57Bl/6 x DBA/2J genetic background and used for SSCs derivation.
Mice were fed with a standard chow diet and acidified water ad libitum. Mice were exposed to a 12 hours light and dark cycle, in rooms at standard temperatures.
Mice were bred and maintained both at the EMBL Mouse Biology Unit, Monterotondo, and at the Scottish Centre for Regenerative Medicine, in Edinburgh, and all experiments were performed under UK Home Office authorisation.

2.1.2 Mice generation using homologous recombination in mouse ESCs

Mice were generated at the EMBL Mouse Biology Unit, Monterotondo, in accordance with the Italian legislation, under license from the Italian Health Ministry.
The construct *pDTA-Tag-Frt-Neo-Frt* used for targeting the Nanos2 locus in ESCs was designed by Dónal O'Carroll and realised with the technical support of Dr. Philip Hublitz, from the EMBL Monterotondo Gene Expression Facility.
For the generation of the *Nanos2*^{TAG} mice, embryos were injected at the morula stages, using the *Nanos2*$^{TAG/+}$ ESCs clones, as described in Chapter 3. This was performed by the EMBL Monterotondo Transgenic Facility.

2.2 Assessment of the fertility of transgenic animals

To validate *Nanos2*^{TAG} males' fertility, both heterozygous and homozygous animals were mated with wild type C57Bl/6 females. 2 months old animals were used to set-up matings and litter size was analysed for the following 3-4 months, at pups' weaning age.

Testicular weight analysis was performed to assess *Nanos2*^{TAG} and *Nanos2*^{Fl} males' fertility. Both heterozygous and homozygous animals were sacrificed when they were 2, 4, 6 months old.

2.3 Genomic DNA extraction for Southern Blotting

A 'Tail digestion buffer' was used for all DNA extraction protocols.
Composition of the tail digestion buffer: 50 mM tris-HCl pH 8; 100 mM EDTA (03609, Sigma); 100 mM NaCl; 1% SDS (L4509, Sigma); 3 mg/ml Proteinase K (3115852001, Sigma Aldrich) to be added fresh.

Tail biopsies or cell suspensions were digested by incubation with 500 µl of Tail Buffer supplied with fresh Proteinase K, at 55 °C, by shaking tubes at 750 rpm overnight. To discard the undigested coat, tubes were centrifuged at maximum speed for 5 minutes at R.T. and supernatants were transferred into a new safe-lock tube (to avoid phenol vapors leakage later on). From this step onwards, DNA isolation was performed under a fume hood. To purify DNA from the organic phase, 1 volume of PCI (Phenol:Chloroform:Isoamyl alcohol) (77617, Sigma) (eg. 500 µl) was added to each tube and samples were mixed by inverting tubes vigorously or vortexing for 20 seconds. Samples were centrifuged at 10 000 rpm, for 5 minutes, at R.T and the upper aqueous phase was collected and transferred into a new tube, without disturbing the white protein interphase. PCI addition and aqueous phase collection was repeated again twice, to achieve high DNA purity. Then, to remove traces of phenol from

the aqueous phase,1:1 volume of Chloroform (32211, SLS) was added. Samples were mixed by inverting tubes and centrifuged as previously described. The aqueous phase was collected and transferred into a new tube. DNA precipitation was achieved by adding 1 volume of Isopropanol and 1:10 Volume of 3 M NaOAc (AM9740, Thermo Fisher). Tubes were inverted vigorously and centrifuged at maximum speed for 5 minutes. After discarding the supernatant, DNA pellets were washed by adding 1 ml of 70% EtOH to remove salt traces. Samples were centrifuged at maximum speed for 5 minutes ad pellets were air-dried for 5 minutes at R.T. DNA pellets were resuspended in 50 -100 μl of ddH20. DNA yield and purity were analysed with a Nanodrop ND-1000 spectrophotometer (Thermo Scientific) and DNA stored at 4 °C.

2.4 Southern blotting

10 μg of genomic DNA, extracted as described before, was digested at 37 °C overnight with 40 U of NheI enzyme (NEB, R0131S). DNA fragments were separated by electrophoresis in a 0.8% Agarose gel and run for 4 hours at 120 Volts. DNA was then transferred to a Hybond Nylon membrane (GE healthcare, RPN203B), using an alkaline solution (1.5 M NaCl, 0.4 M NaOH). The membrane was rinsed twice, dried for 20 minutes at 37 °C, and cross-linked with UV-light at 150 mJ/cm^2.

The NANOS2-DNA probe used was complementary to a region enclosed between the last two NheI sites, in *Nanos2* locus. 30 ng of this probe were marked by using the Random primers labelling kit (18187-013, Invitrogen) with α[32P]-dGTP. The membrane was incubated at 65 °C for 2 hours with the hybridization buffer (0.5M Na-Phosphate buffer pH 7.2; 1mM EDTA, (Sigma, 03609); 3% BSA;
5% SDS (Sigma, L4509)), and then with the labelled probe overnight. The membrane was washed for 20 minutes at 65°C, using the pre-heated washing solution (40mM Na-Phosphate buffer pH 7.2; 1mM EDTA; 5% SDS). These

washes were repeated for three times. The membrane was then exposed for 24 hours to a phosphoscreen, which was scanned later using a Fluorescent Image Analysis System (Fujifilm, FLA-5100).

2.5 PCR genotyping

2.5.1 Genomic DNA extraction for PCR Genotyping

Tails or ear biopsies were digested with 400 μl TAIL digestion buffer supplied with fresh Proteinase K for 3 hours (or overnight), at 55 °C, 750 rpm. To remove the undigested coat, tubes were centrifuged at maximum speed for 5 minutes and supernatants containing lysed cells were transferred into a new tube. To precipitate DNA out of solution, a 1:1 volume of isopropanol was added to each sample and tubes were vortexed before centrifuging them at maximum speed for at least 15 minutes. Supernatants were decanted and the DNA pellet were washed with 600 μl of 70% ETOH to remove residual salts. Sample were centrifuged as previously explained and supernatants decanted. Remaining Ethanol was removed by another centrifugation and by air-drying DNA pellets for approximately 10 minutes. DNA was resuspended in 40 μl of ddH20 and 1 μl was used as a template in a 25 μl PCR volume, for genotyping.

2.5.2 Genotyping protocols

All PCRs were resolved by electrophoresis in a 3% Agarose (A9339, Sigma) gel, containing 5% EtBr (09-0617 Sigma Aldrich), in 1X TAE buffer (40 mM Tris, 20 mM acetate, 1 mM EDTA), run at 200 Volts for 30-40 minutes.

PCR mix:

2.50 μl	10X DreamTaq Green Buffer (B71, Life technologies)
0.25 μl	10 mM dNTPs (10319879, Thermo Scientific)

0.25 µl 20 µM primer mix

0.40 µl Taq Polymerase (EMBL Heidelberg Facility)

20.60 µl H_2O

Genotyping protocol for *Nanos2^{TAG}* allele

Primers:

"FW1_N2" : 5'- AACCTGGGGAATAACCTGCT-3'

"FW2_N2" : 5'- TGCTGCTGAATAAAGCGTTG-3'

"RV_N2" : 5'- TCCCAGTCAGACGACTTGTG-3'

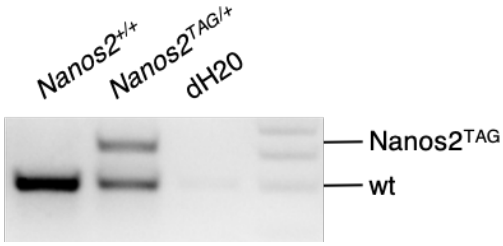
PCR program:

Step	TEMERATURE (°C)	Time (s)	CYCLES
1. Denaturation	95	180	1 X
2. Denaturation	95	30	34 X
3. Annealing	63.3	30	
4. Extension	72	35	
5. Final extension	72	300	1 X
6.	12	Hold	-

Expected band's length:

Wt: 212 bp

Nanos2^{TAG}: 353 bp



Genotyping protocol for FLP allele

Primers:

"DOC_FLP1" : 5'- CCCATTCCATGCGGGGTATCG -3'

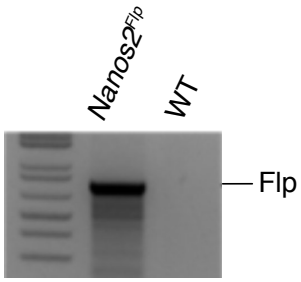
"DOC_FLP2" : 5'- GCATCTGGGAGATCACTGAC/G -3'

PCR programme:

Step	TEMERATURE (°C)	Time (s)	CYCLES
1. Denaturation	90	240	1 X
2. Denaturation	90	60	33 X
3. Annealing	65	60	
4. Extension	72	60	
5. Final extension	72	600	1 X
6.	15	Hold	-

Expected band's length:

Nanos2^{Flp}: 650 bp



2.6 Histology

Two-, four- and six-months old mice, from *Nanos2*^{CTL}, *Nanos2*^{TAG} alleles, were sacrificed and testes were dissected to examine the histology of seminiferous tubules. Four animals per genotype were analysed. Testes were fixed in Bouin's solution (16045-1, Polysciences) and embedded in paraffin. Testes were sectioned with a Microtome to obtain 4-µm-thick slices, every ~300-400 *µ*m, to have ~15 sections for testis. Testes slices were obtained from the beginning, a quarter and half of the testis longitudinal length. Slices were placed on a glass slide and these were H&E stained, according to standard protocols. Images of tubules' sections were acquired with Zeiss Axio Scan Slide Scanner and the presence of spermatogenic cells was evaluated by analysing images with the Zen Airyscan and Fiji ImageJ softwares.

2.7 Whole-mount immunofluorescence of seminiferous tubules

Adult mice were sacrificed, and testes were dissected. The genotype of each mouse was confirmed by taking a tail biopsy for DNA extraction and PCR genotyping. After weighing testes, their tunica albuginea was removed, and seminiferous tubules were gently spread with tweezers. Testis were rinsed in DMEM (41965039, Life technologies) medium enriched with 1 mM Na-Pyruvate (11360070, Life technologies) and 1.5 mM Sodium DL-Lactate (L4263, Sigma). In order to release tubules from the connective tissue, testes were digested with 0.5 mg/ml Collagenase Type XI, (C7657, Sigma-Aldrich) in a water-bath at 32 °C for 7 mins, at 700 rpm. After removing the medium and washing them with PBS, tubules were fixed by covering them with 4% PFA and incubating for them 4 hours, on a rocker, at 4°C. PFA was washed away with PBS. Tubules were transferred into a 48-well plate and analysed under a stereomicroscope, in order to select the best tubules (long and intact). PFA was quenched with 1 M Glycine for 30 min at R.T., rocking. Tissue permeabilisation and blocking were performed by incubating tubules with PBS/0.5% TritonX-100 / 5% serum / 0.5% BSA (the serum used depended on the animal where secondary antibodies were made). Tubules were incubated with the appropriate primary antibody in 200 ul volume of 0.5% BSA, PBS, at R.T., over night. Primary antibodies used are: α-GFP, 1:500 (GFP-1010, Aves), α-GFRα1, 1:50 (GT15004, Neuromics), and α-c-kit, 1:250 (AF1356-SP, R&D Systems). Antibody excess was washed away with PBS, for three times. Tubules were stained by incubating with Alexafluor-488, or -546 conjugated secondary antibodies, in PBS, 0.5% BSA, 0.1% Triton X-100, at R.T., for 1 hour. Tubules were washed for three times with PBS. Nuclei were counterstained with a solution of 1:1000 4',6-diamidino-2-phenylindole (DAPI) (D1306, Life Technologies) (5 µg/ml) in PBS, for 10 minutes, at R.T. After washing the excess of DAPI with PBS, tubules were mounted in ~50 μl of ProLong Antifade Mountant (P36930, Life Technologies) on a glass slide (2800080, Leica). Twenty images per animals were acquired by using a

Confocal microscope, by taking Z-stacks, and with Z-stepsize set at 0.34 μm. Images were analysed with Fiji ImageJ.

2.8 Adult spermatogonia preparation for FACS

Adult mice were sacrificed, and testes were collected. After removing the tunica albuginea, testes were rinsed in DMEM, 1 mM sodium pyruvate, 1.5 mM sodium lactate and digested with 0.5 mg/ml Collagenase Type XI, for 7 minutes, at 32 °C, at 700 rpm. Tubules were then digested with 0.05% Trypsin-0.05 mg/ml DNase I (DN-25; Sigma-Aldrich), to obtain a single-cell suspension. Cells were filtered using a 70-µm strainer (14–959-49A; Falcon) and pelleted at 1,000 rpm, for 5 minutes. For each sample, approximately 5 million cells were stained using the appropriate antibodies diluted in PBS 3%/FBS/0.01% sodium azide in ice. Antibodies used were either fluorophore-conjugated: CD117 (c-kit) (PE-Cy7 15360940), CD9-APC (Thermo Scientific, 15360570), or primary antibodies: CD45 (eBioscience, clone A20), CD51 (BioLegend, clone RMV-7). Cells were incubated with antibodies on ice and tubes were mixed every 5 minutes. For samples stained with CD45 and CD51, cells were washed with PBS-FCS and incubated with Qdot 605 Streptavidin Conjugate (Q10101MP, Life technologies) for 30 minutes, on ice. Cells were filtered, stained with SYTOX Blue (S34857, Thermo Scientific), and then analysed by FACS.

2.9 Cells culture

All cells were grown at 37 °C. HEK293t cells were grown in a 5% CO_2 incubator whereas MEFs and SSCs were cultured in 7.5 % CO_2 incubator.

2.9.1 HEK293t culture

HEK293t cells were cultured in the following media: GMEM (12-739F, Lonza), 10% fetal bovine serum (FCS) (Life technologies), 2 mM L-glutamine (Invitrogen), 100 units/100 μg Penicillin/Streptomycin (P4333, Sigma), 1X Non-essential amino acids (11140050, ThermoFisher Scientific), 0.1 mM 2-Mercaptoethanol (31350-010, ThermoFisher Scientific). HEK293t cells were cultures in 10 cm (SIAL0599, Sigma) or 15 cm plates (430167, Corning) and passaged when they were 70-80% confluent, to a ratio of 1:10 (generally). Media was aspirated gently and cells were washed with PBS (D8537, Sigma). Cells were detached by resuspending them with pre-warmed HEK media and transferred into new plates.

2.9.2 Transfection of HEK293t cells

HEK293t cells were re-plated in 15 cm plates one day before transfection, in order to be 70% confluent on the day of transfection. Fresh medium was added to cells before transfection. Transfection was performed as described in the Lipofectamine manufacturer's protocol. 15 μg of plasmid DNA and Lipofectamine 2000 (11668030, Invitrogen) were first diluted in Opti-MEM medium (31985062, ThermoFisher Scientific) and then mixed gently in one tube. DNA/lipofectamine complexes were added dropwise to cells and plates were placed back in the incubator and kept there for 24 hours. Transfected cells were visualised with a fluorescence microscope.

2.9.3 Derivation and passaging of MEFs

After setting-up matings, females were examined for the presence of vaginal plugs. Matings were separated after detection of plugs and females were sacrificed 13.5 days later (plug day counts as 0.5 day), in order to harvest embryos. Embryos were dissected, placed in a 10cm plate and covered with

cold PBS. In order to isolate MEFs from embryos, desidua, placental and maternal tissues were removed. Embryos' heads were cut away and all innards (blood tissues, liver etcetera) were removed by eviscerating the embryos' body. To dissociate the remaining embryos' tissues, these were collected in a 10 ml syringe. By adding 5 ml of MEF media into the syringe, embryos were passed through a 18-gauge needle for three times. This operation was repeated using a 21-gauge needle. The homogenized tissues were kept in the 10 cm plates and media was added to reach a volume of 10 ml (this is passage 0, P0). Plates were placed in the incubator and cell were grown for three days to reach 70-80% confluency. Cells were passaged in the following way: first, they were rinsed with PBS, digested with 0.05% Trypsin (25300054, Life technologies) for three minutes at 37 °C and then they were collected in a tube. Clumps were removed by sedimentation. Cells in the supernatant were pelleted and resuspended in appropriate volume of MEF media. Each 10 cm plate was passaged into a 15 cm plate (passage 1, P1). Cells were grown for two days to reach 80-90% confluency and passaged as previously described (without the clumps' sedimentation step) to a ratio of 1:3, in other 15 cm plates (passage 2, P2). Cells were grown for two days to reach 80-90% confluency and they were either frozen or passaged again to a ratio of 1:3 (passage 3, P3). Once MEFs reached 80% of confluency, they were irradiated to induce DNA damage and subsequent cell cycle arrest. To irradiate MEFs, cells at P3 were trypsinised, resuspended in MEF media and collected in a 50 ml tube. Tubes were placed in a RS 2000 X-ray irradiator (Rad Source), and the X-ray dose was set at 20 Grays/minutes. Cells were placed back in the incubator for ~10 minutes, pelleted by centrifugation and resuspended in MEF media. Inactivated MEFs were used as feeders for SSCs culture. Thus, inactivated MEFs were plated in new plates and the following day SSCs were plated on top of them. Alternatively, when MEFs were not immediately used, they were resuspended in freezing media and stored in Liquid nitrogen.

MEF media composition: DMEM, 12.5% FCS, 2 mM L-glutamine, 1X non-essential amino acids, 100 U/ml penicillin/streptomycin and 100 µM β-mercaptoethanol (Gibco).

2.9.4 Derivation and maintenance of SSC lines

SSCs derivation and maintenance protocol was adapted from (Kanatsu-Shinohara et al., 2003).

DBA/2 mice were bred with BL6N mice and F1 or F2 hybrid pups were used for SSCs derivation, when they were ~7 days old.

After sacrificing the pups, a tail biopsy was taken to confirm the genotyping. To isolate testes, the abdomen of each pup was opened, and testes transferred into 1 drop of cold HBSS buffer (14170-088) on a plate. The tunica albuginea was removed, and testes were washed with HBSS buffer twice. To isolate germ cells, testes were digested with 600 μl of Collagenase (1mg/ml in HBSS) in a 15ml falcon and incubated for 10-12 minutes at 37°C in a water bath, by gently shaking every 5 minutes. Tubules were centrifuged at 800 rpm for 3 minutes. Sedimented tubules were washed in HBSS and resuspended in 800 μl of 0.25% Trypsin (25200056, Life technologies), supplemented by 200 μl of DNase (7 mg/ml in HBSS). Tubules were incubated for 10 minutes at 37°C. Trypsin was blocked by adding 5 ml of MEF media and tubules were dissociated by pipetting 10 times with a p200 gilson pipette. The suspension was filtered with 0.45 μm filters and centrifuged at at 800 rpm for 5 minutes. Cells were resupended in 1 ml of SSCs media and counted with a Nexcelom (Bioscience) automated cell counter. SSCs were plated to have a final concentration of $3x10^5$ cell/ml, in a well of gelatin coated 12 well-plates. Plates were incubated for 12 hours at 37°C. By collecting floating cells only, SSCs were transferred to new 12-well plates, thus depleting somatic cells. Media was changed every 3 days and after observing the first SSCs colonies (~6-10

days later), SSCs cultures were split 1:1 (Passage1). SSCs were cultured for additional 4-5 days, or until colonies of 10-20 cells became visible. At that stage SSCs were depleted from somatic cells.

For the SSCs depletion from somatic cells, SSCs medium from SSCs cultures was collected in tubes. SSCs were washed twice with PBS and then they were trypsinised with 200 μl of 0.25% Trypsin for each well, for 3 minutes at 37°C. Trypsinisation was blocked with MEFs media and tubes were centrifuged. SSCs were resuspended in the SSCs media previously collected and plated on gelatin-coated plates. SSCs were left in the incubator for 30 minutes-1 hour, the time necessary to allow deposition of somatic cells on the bottom of the plates. Floating SSCs were collected and counted. 8 x 10^5 cells were plated in a well of a 6-well plate, which was previously seeded with 3-4 x 10^5 inactivated MEFs. This constituted passage number 2 of the derivation protocol.

For SSCs maintenance, SSCs media was changed every 3 days and SSCs passaged every 7-8 days, 1:2 or 1:3. One day before passaging them, inactivated MEFs were thawed and plated in the appropriate plates.
SSCs were washed twice with PBS and trypsinised as described before. As soon as MEFs media was added to block trypsinisation, MEFs clumps (white) were collected from the plates with a Gilson p1000 set to 1 ml and transferred to 15 ml tubes. The remaining cells, enriched for SSCs, were collected and added to the same tubes, without resuspending. Tubes were left at R.T. for 5 minutes, to allow MEF sedimentation. Supernatants were transferred to new tubes and centrifuged. MEFs clumps in the bottom of the tubes were discarded. SSCs pellets were resuspended in the appropriate amount of pre-warmed SSCs media and plated on MEF-coated plates.

Reagents for SSCs cultures:

HBSS, Invitrogen (14170-088); DNAse, Sigma (DN25); Collagenase type I, Sigma (17100-017); Trypsin-EDTA 0.25%, Invitrogen (15050-065); Ultrapure water with 0.1% gelatin, Millipore (ES-006-B).

Spermatogonial Stem Cell media:
Stem Pro-34 SFM medium with 2,6 % Supplements, Invitrogen (10639-011); 6 mg/ml D(+)Glucose, Sigma (G7021); 25 μg/ml Insulin, Sigma (I5500); 5 ml/ml BSA, MP-Biomedicals (810661); 100 μg/ml Transferrin, Sigma (T1147); 30 nM Sodium Selenite, Sigma (S5261); 1X MEM Vitamin solution, Invitrogen (11120-0370); 10 μg/ml D-Biotin, Sigma (B4501); 60 ng/ml Progesterone, Sigma (P8783); 30 ng/ml β-estradiol, Sigma (E2758); 1X Sodium Pyruvate, Sigma (P2256-5G); 60 μM Putrescine, Sigma (P7505); 1% L-Glutamine, Invitrogen (25030-024); 5.7 10^{-7} M β-mercaptoethanol, Sigma (M3148); 1 μl/ml Lactate, Sigma (L4263); 1% FBS, Millipore (ES-009-B); 20 ng/ml mEGF, B&D (354001); 10 ng/ml, hbFGF, B&D (354060); 10 ng/ml rat-GDNF, R&D System (512 GF); 20 ng/ml Recombinant LIF; 1% PenStrep, Invitrogen (15070-063).

2.9.5 Cell freezing

Cells were frozen when they reached at least 80% confluency. Cells from one 6-wells plate or one 10 cm plate were frozen in 1 cryotube vial. Cells were trypsinised, collected by centrifugation and resuspended in freezing media, composed of 50% FCS (or 50% HyClone FBS (12379802, Fisher Scientific) for SSCs), 40% cell specific media (HEK, MEF or SSC media), 10% DMSO (D2650, Sigma). Vials were quickly moved into home-made freezing chambers to allow slow freezing. Chambers were made by a polystyrene rack, where cryotubes were placed, covered by paper towels and by another polystyrene rack, placed on top as a lid. The two racks were taped tightly and moved to the

-80C. After two days tubes were transferred to the liquid nitrogen, for long term storage.

2.9.6 Cell thawing

All cells types were thawed rapidly in a 37 °C water-bath and taken out when tubes were still cold. Each vial was resuspended in 10 ml of either HEK293t media (for HEK293t cells) or MEF media (for both MEFs and SSCs). Cells were pelleted by centrifuging tubes at 1000 rpm (for HEK293t and MEFs cells) or 2000 rpm (SSCs), for 5 minutes. Pellets were resuspended in the appropriate media, before plating cells in treated-plates.

2.10 Immunoprecipitation followed by mass spectrometry (IP-MS)

Samples for IP-MS were prepared as described in (Much et al., 2016). Samples were analysed by nanoscale liquid chromatography coupled to tandem mass spectrometry. Peptides were separated on an EasySpray (Thermo Fisher) 50 cm column coupled to an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher). Data were analysed by using MaxQuant version 1.6.1.0., and label-free quantitation (LFQ) was performed by using the MaxQuant LFQ algorithm (Cox et al., 2014). Peptides were analysed using the Uniprot database, and common contaminants (e.g. trypsin, keratins, etc.) were removed during the Perseus analysis (Cox et al., 2014; The UniProt Consortium, 2017; Tyanova et al., 2016; Hubner et al., 2010). For visualisation, LFQ intensities obtained from MaxQuant analysis were imported into Perseus version 1.6.0.2 (Tyanova et al., 2016) and processed as described (Hubner et al., 2010).

2.11 RNA isolation

SSCs were MEFs-depleted and sorted by FACS according to size and shape. Approximately 1.5 x 10^6 SSCs were sorted for each sample and resuspended in cold PBS-2%FCS. These were pelleted by centrifuging at 2000 rpm, for 5 minutes at 4 °C, by using a centrifuge with swinging-bucket rotor. Supernatants were gently discarded, and cellular pellets were resuspended in 1 ml of QIAzol Lysis Reagent (79306, Qiagen). Cell lysis was performed by incubating samples at R.T. for 5 minutes. 200 μl of Chloroform was added to each sample and mixed by pipetting. Tubes were incubated at R.T. for 2-10 minutes and centrifuged for 10 minutes at maximum speed, at 4 °C. Only the aqueous phase was transferred into new tubes, without dragging the white interphase. RNA was precipitated by adding 400 μl of Isopropanol and vortexing. Also, a nucleic acid carrier (1 μl of 5 μg/μl Linear acrylamide (AM9520, Life technologies)) was added in order to visualise the pellet. Samples were kept at R.T. for 10 minutes, and then centrifuged for 10 minutes or up to 1 hour. Supernatants were discarded, and the white pellet containing RNA was washed with 400 μl of 70% EtOH. Tubes were centrifuged at 7500 rpm for 5 minutes and pellets were air-dried and then resuspended in 10-25 μl of DEPC nuclease free water (AM9916, Ambion).

2.12 RNA quality controls

The quality and integrity of RNA samples was assessed with the High Sensitivity RNA ScreenTape (5067- 5579, Agilent Technologies). RNA samples concentration was measured with Qubit RNA HS Assay Kit (Q32852, Life Technologies).

2.13 mRNA-seq

1.2 μg of total RNA (isolated as previously described) was treated with DNAse I, in 10X Buffer (AMPD1, Sigma) for 30 minutes at 37 °C. RNA was purified from components of the DNAse I reaction using Rneasy MinElute columns (74204, Quiagen) suitable for the purification of mammalian mRNAs (longer than 200 nt). The elution was done in 20 μl of nuclease free water. All the purified RNA was used for Ribo-depletion with the RiboCop kit (see below) and RNA eluted in 12 μl of Elution Buffer. 9 μl of Ribo-depleted RNA was initially used to start Library generation with with SENSE Total RNA-Seq Library Prep Kit (include RiboCop), (SKU: 042.24, Lexogen). After reverse transcription and second strand synthesis, the double-stranded libraries were purified with 13 μl of Bead Diluent and 27 μl of Purification Solution, in order to select libraries with mean size of 386 bp and insert size of 264 nt. Libraries were amplified with 18 cycles of PCR.

2.14 CRAC in HEK293t cells and SSC lines

The *pyCAG-Tag-Nanos2* plasmid for the generation of *Nanos2*^{OE} HEK293t cells was cloned by using the pyCAG-vector (Chambers et al., 2003), which was a gift from the Chambers laboratory.

The DNA corresponding to Tag-Nanos2 was amplified from genomic DNA obtained from a biopsy of *Nanos2*^{TAG} mice, by performing a PCR with the following primers:

FW_TAG-Nanos2_HEK: 5'-TTTctcgagATGGGTAAGCCTATCCCTAACC-3'
RV_TAG-Nanos2_HEK: 5'-TTTgcggccgcTTATCGCTTGACTCTGCGAC-3'

Primers for Sanger Sequencing:

FW 941 Nos: 5'-AGGAGGGTGAGGTAGCTGAG-3'
FW 586 GFP: 5'-ATCATGGCCGACAAGCAGAA-3'
FW 77 N PreSci: 5'-ATCTGCTGGAGGTGCTGTTC-3'

UV-Cross-linking and cell lysate preparation

HEK293t cells were grown in two 150 cm^2 plates and were transiently transfected (as explained before) to obtain *Nanos2*^{OE} cells. As a control, other two plates of cells were grown and Mock-transfected. Two days after transfection, cell media was discarded and cells were washed twice in PBS (14190094, Life technologies).

Nanos2^{CTL} and *Nanos2*^{TAG} SSCs were grown on feeders in eight 150 cm^2 plates each genotype. Media was aspirated and cells were washed twice in PBS.

At this stage, both HEK293t and SSCs were covered with a minimum amount of PBS (~ 3 ml) and plates were quickly placed on ice and irradiated at 254 nm (Bulbs, 400073, Agilent), with energy set at 400 mJ/cm^2, by using a Stratagene Stratalinker 1800. By keeping plates on ice, cells were scraped off and lysed with 2 ml of Lysis Buffer (supplied with fresh complete EDTA-free) per plate. Lysates were pooled into low binding tubes and left on ice for additional 10 minutes. Lysates were centrifuged for 20 minutes at maximum speed, at 4 °C. Clear supernatants (cytosolic fraction) were collected and nuclear pellets discarded. When clear lysates were not processed immediately for IP, they were snap-frozen in liquid Nitrogen and stored at – 80 °C. Inputs were also collected from the clear lysate for quality control / Western Blot (WB); they were boiled in Sample Buffer (NP0007, Life technologies) for 5 minutes and frozen.

Purification of Protein-RNA complexes

1st step: protein IP

Anti-V5 (R960-CUS, Invitrogen) antibody was coupled with Dynabeads (11206D, Life technologies), for 15 minutes at R.T on a rotating wheel. Anti-V5-Beads were isolated by using a magnetic rack and incubated with protein lysates for 2 hours, on a rotating wheel, at 4 °C. Beads were separated from

the flow-through (FT) on a magnetic rack. FT were collected for WB analysis. Beads were washed twice with 5 ml of TNM 600 buffer and once with 5 ml of TNM 100 buffer. Beads were resuspended in 1 ml of TNM 100 buffer without protease inhibitors.

2nd step: protein cleavage

Resuspended beads were incubated with 20 µl of 3C-GST Protease (2 mg/ml, EMBL) for 2 hours at 4 °C. Beads were separated from the supernatant on a magnetic rack and an aliquot from each fraction was saved for WB.
Supernatants were incubated with 0.5 U of RNAce-It (400720, Agilent), for 5 minutes at 25 °C. Subsequently, a mixture of 0.5 g guanidine-HCl, 300 mM NaCl, 10 mM imidazole was added to the samples and vortexed.

3rd step: Ni-NTA binding and denaturing washes with GuHCl

In the meanwhile, 50 μl of nickel agarose beads (30410, Qiagen) were washed twice with 1 ml Ni-WBI, (2000 g, 1 minute, at 4 °C). Beads were incubated with samples overnight on a rotating wheel, at 4 °C. Beads' flow throughs (FT) were collected for protein precipitation and WB. 15 μl of FT was mixed with 185 μl PNK buffer and six volumes of pre cold acetone (32201, Fisher Scientific). The mixture was incubated on dry ice for few hours and then centrifuged at maximum speed for 20 minutes, at 4 °C. Supernatants were discarded, pellets were dried and processed for WB.

Beads were washed three times with 1.5 ml Ni-WB I, incubated for 15 minutes each wash at 4°C, with nutation. Afterwards, beads were washed twice with 1 ml of Ni-WB II and then three times with 0.75 ml of PNK buffer. The last wash was performed after transferring the beads onto microbio spin columns (69725, Sigma) For the phosphates treatment, the following 80 μl mix was added to the columns: 1X PNK buffer, 8 U TSAP (M9910, Promega), 80 U RNasin (N2111, Promega), and samples were incubated for 30 minutes at 37 °C. TSAP enzyme was inactivated by adding 400 μl of WB-I buffer to the columns and washed three times with 400 μl of PNK buffer to remove

guanidine-HCl. A pre-adenylated linker was ligated to the 3' end of RNAs by adding 80 μl of 1X PNK buffer, 10 μM miRCat-33linker (/5rApp/NAGATCGGAAGAGCACACGTCTG/3ddC/), T4 RNA ligase II truncated (M0242L, NEB), 40 U of RNasin, 10% PEG. Samples were incubated at 22 °C, for 4 hours, at 300 rpm and subsequently washed once with 0.5 ml of Ni-WBI, twice with 0.75 ml of Ni-WBII and twice with 0.75 ml of PNK buffer. RNA was radiolabeled by applying to samples 80 μl of the following labeling mixture: 40 μCi ^{32}P-γATP (NEG502Z250UC, Perkin Elmer), 1X PNK buffer, ^{32}P -γ-ATP (NEG502Z250UC, Perkin Elmer), 20 U of T4 polynucleotide kinase PNK (M0201S, NEB), 40 U of RNasin and incubating columns for 30 minutes at 37 °C. The labeling reaction was stopped by adding 1 mM of cold rATP (E6011, Promega) and incubating samples for 15 minutes at 37 °C. 5' linkers were ligated to RNA by incubating samples with 10 pmol of 5' linkers and 40 U of T4 RNA ligase I (M0294L, NEB), at 16 °C overnight. Samples were washed once with 1 ml of Ni-WBI and three times with 1 ml of PNK buffer. RNPs were eluted from the columns with 300 μl of Elution Buffer, for 5 minutes at 23 °C, 500 rpm and precipitated by adding 2 μg of BSA, 1.5 ml of cold acetone and incubating over night at -20 °C. Radiolabeled RNPs were resolved on a 4-12% Bis-Tris NuPAGE gel (NP0335BOX, Invitrogen), using NuPAGE Running Buffer (NP0001, Life technologies) and transferred with NuPAGE Transfer Buffer (NP00061) to a nitrocellulose membrane at 70 Volts, for 1.5 hours, on ice. RNPs were detected by autoradiography and their correct molecular size was identified by overlapping the film with the membrane showing the protein ladder (Kaleidoscope, 1610375, Bio-Rad). RNPs were extracted by cutting the membrane and incubating the slices with 100 ug of proteinase K in WB-II supplied with 1% SDS and 5 mM EDTA, for 2 hours at 55 °C. RNA was extracted by adding sodium acetate (AM9740, Life technologies) and PCI (15865408, Fisher Scientific). Tubes were vortexed, centrifuged for 5 minutes, 20000 g, at 4 °C. The aqueous phase was collected, and RNA was precipitated overnight at -20 °C, by adding 1 μl GlycoBlue (AM9515, Thermo fisher) and 1.1 ml of EtOH. The RNA pellet was washed

with 70 % EtOH, centrifuged and air-dried. RNA was resuspended in 13 μl of the Reverse transcription mixture containing the mirCat-33 primer and dNTPs and incubated for 3 minutes at 80 °C. Samples were snap-chilled on ice and kept there for 5 minutes. Subsequently, First-strand buffer, DTT and RNasin were added and incubated at 50 °C for 3 minutes. 1 ul of Superscript III (18080044, Life technologies) was added and incubated for 1 hour at 50 °C. The enzyme was inactivated by incubation at 65 °C for 15 minutes. To release the cDNA, samples were incubated with RNaseH and incubated for 30 minutes at 37 °C. The reverse transcription product was aliquoted and used for PCR. 5 ul of cDNA was used for 50 ul PCR containing TaKaRa LA Taq (RR002M). PCR were pooled and purified from PCR reaction components by precipitating DNA with sodium acetate, 100% EtOH and freezing at -80 C for 2 hours. Samples were centrifuged at max speed for 30 minutes at 4 °C. DNA pellets were washed with 70% EtOH and resuspended in dH20. PCR products were resuspended in Loading Dye and resolved in 3% MetaPhor agarose gel (Lonza, LZ50180) and PCR fragments with inserts size between 10-50 bp (total length 170-200 bp were extracted, gel purified with MinElute Gel extraction kit (28604, Qiagen) and resuspended in dH20 for deep sequencing.

Lysis	Per 50 ml
50 mM Tris-HCl pH 7.8	2.5 ml 1M stock
0.1 M NaCl	1,67 ml 3M stock
1 % NP-40	5 ml 10 % stock
5 mM $MgCl_2$	0.25 ml 1 M stock
10% glycerol	10 ml 50% stock
5 mM B-ME (2860 x)	Added fresh

TNM100	Per 50 ml
50 mM Tris-HCl pH 7.8	2.5 ml 1M stock
0.1 M NaCl	1,67 ml 3M stock
0.1 % NP-40	0.5 ml 10 % stock
5 mM $MgCl_2$	0.25 ml 1 M stock
5 mM B-ME (2860 x)	Added fresh

TNM600	Per 50 ml
50 mM Tris-HCl pH 7.8	2.5 ml 1M stock
0.6 M NaCl	10 ml 3M stock
0.1 % NP-40	0.5 ml 10 % stock
1.5 mM $MgCl_2$	75 μl 1M stock
5 mM B-ME (2860 x)	Added fresh

PNK buffer	Per 50 ml
50 mM Tris-HCl pH 7.8	2.5 ml 1M stock
50 mM NaCl	0.83 ml 3 M stock
0.1 % NP-40	0.5 ml 10 % stock
1.5 mM $MgCl_2$	75 μl 1 M stock
5 mM B-ME (2860 x)	Added fresh

WBI (wash buffer I for Ni-NTA)	Per 50 ml
50 mM Tris-HCl pH 7.8	2.5 ml 1M stock
300 mM NaCl	5 ml 3M stock
0.1 % NP40	0.5 ml 10 % stock
10 mM Imidazol	0.2 ml 2.5M stock
1.5 mM $MgCl_2$	75 μl 1 M stock
6 M GuHCl	28.66 g
5 mM B-ME (2860 x)	Added fresh

WBII (wash buffer II for Ni-NTA)	Per 50 ml
50 mM Tris-HCl pH 7.8	2.5 ml 1M stock
300 mM NaCl	5 ml 3M stock
0.1 % NP40	0.5 ml 10 % stock

10 mM Imidazol	0.2 ml 2.5M stock
1.5 mM $MgCl_2$	75 μl 1 M stock
5 mM B-ME (2860 x)	Added fresh

EB (Ni-NTA elution buffer)	Per 50 ml
50 mM Tris-HCl pH 7.8	2.5 ml 1M stock
50 mM NaCl	0,87 ml 3M stock
0.1 % NP40	0.5 ml 10 % stock
150 mM Imidazol	3 ml 2.5M stock
5 mM B-ME (2860 x)	Added fresh

PK digestion buffer (for RNA extraction)	Per 10 ml
50 mM Tris-HCl pH 7.8	1M stock
50 mM NaCl	3M stock
SDS	20% stock
EDTA	3M stock
0.1 % NP40	10 % stock
150 mM Imidazol	2.5M stock
5 mM B-ME (2860 x)	3,4 ul Added fresh

5' linkers used for the CRAC in SSCs:

5' linker name	**5' linker sequence**
L5Aa (TA)	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrUrArArGrC-OH-3'
L5Ad (CGCTT)	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrCrGrCrUrUrArGrC-OH-3'
L5Cb (TGGAGC)	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrUrGrGrArGrCrArGrC-OH-3'
L5Cc (ACTC)	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrArCrUrCrArGrC-OH-3'

L5Cd (GACTT)	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrGrArCrUrUrArGrC-OH-3'

2.15 CRAC analysis

CRAC analysis was performed by Dr. Turowski, by following the principles described in the PyCRAC software (https://bitbucket.org/sgrann/pycrac).
Sequencing reads were pre-processed as follows: reads were demultiplexed removing in-line barcodes and random barcodes (using the pyBarcodeFilter.py script from pyCRAC package v3.0 (Webb et al., 2014)). Then, they were collapsed to remove PCR duplicates by using fastx_collapser v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/). The 3' adapter was removed by using flexbar v3.4.0 (Dodt et al., 2012), which was also used for quality filtering. Pre-processed reads were aligned to the mouse genomic database (mm10) using two different alignment programs, STAR v.2.5.3a (Dobin et al., 2013) and Salmon v.0.13.1 quasi-quantification (Patro et al., 2017). Reads were also aligned to the mouse transcriptome database (Biomart (Smedley et al., 2015)) by using Novoalign v2.07.00 (http://www.novocraft.com).
Peaks were called using the SciPy python library (function argrelextrema) to find local extreme or by simply calling signal extrema to localise the top peak of each transcript (FDR < 0.001). Nanos2 targets were initially ranked according to their total peak score (in rpm, reads per million), which was calculated by summing the intensities of one or more unique peaks present in a given transcript. For each transcript, the total peak score was calculated from two CRAC replicates and the average value for used eventually to generate the final CRAC list. CRAC peaks were visualised with the IGV (Integrative Genomics Viewer) (Robinson et al., 2011).

2.15.1 CRAC normalisation using mRNA-seq

The bioinformatic analysis was performed by Dr. Louie van de Lagemaat.

FPKM values obtained from mRNA-seq data were calculated as the mean between two biological replicates (*Nanos2*$^{TAG/+}$ and *Nanos2*$^{TAG/TAG}$). Only genes containing both CRAC peaks and detectable gene expression levels were considered for the analysis: non-expressed genes (FPKM = 0) were eliminated in this comparison. Normalisation was performed by dividing the FPKM mean value of each gene with the corresponding CRAC peak score. The resulting values were used to rank again the CRAC peaks in a new list, which was called 'normalised CRAC list'. The correlation between CRAC peak scores and FPKM values was calculated with the "CORREL" function in Excel.

2.15.2 Motif analysis

De novo motif analysis was performed by using MEME-ChIP (Bailey et al., 2009). As input for MEME-ChIP we used the normalised NANOS2-CRAC peaks, containing deletions (deletions were previously detected in CRAC peaks, as described in the PyCRAC software). Sequences of 100 bp were centered on the peak summit, and on deletions. We scanned only one strand of the CRAC sequences and looked for motifs with E-value < 0.05, between 6-15 bp, with 0-1 or 1 occurrence per sequence. As a background, we shuffled the 3' UTRs of NANOS2 CRAC peaks by using MEME-ChIP. ATTRACT (Giudice et al., 2016) was used to look for known motifs.

2.15.3 Gene ontology (GO) analysis

GO analysis was performed by using the g:Profiler software (Reimand et al., 2019). The input for the GO analysis was a ranked list of normalised NANOS2-CRAC peaks. Biological terms were ranked according to the adjusted P-value.

2.15.4 Gene clusters analysis

This analysis was performed by Dr. Louie van de Lagemaat. mRNA-seq data from ex-vivo sorted spermatogonia were used as input for the analysis. To detect changes in the gene expression levels across the three populations of spermatogonia, genes were clustered using the Markov clustering algorithm (MCL) (Freeman et al., 2007). Normalisation of the reads was performed by using DESeq2 (Love et al., 2014). To evaluate the enrichment of NANOS2 CRAC targets across the different gene clusters, we used the hypergeometric test.

2.16 SLAM-seq

SLAM-seq was performed by following the manifacturers' instructions in the SLAMseq Kinetics Kit – Catabolic Kinetics Module, 24 preps, SKU: 062.24, Lexogen.

All experiments were performed in the dark, or in the presence of a red-light lamp. After determining the optimal concentration of s^4U (100 μM) with the cell viability tritation assay, SSCs media containing s^4U was used to label SSCs for 24 hours. s^4U was supplied also three hours before time 0. At time 0, s^4U-media was replaced with fresh SSCs media containing an excess of unlabeled uridine. SSCs were collected at different time-points: 0, 30 minutes, 1 hour, 2 hours, 9 hours, and 24 hours. SSCs were collected by trypsinisation, and MEFs depletion (as described in the SSCs culture section). Additionally, SSCs were FACS sorted to exclude dead cells and MEFs contaminants. Two biological replicates, constituted by *Nanos2*$^{TAG/+}$ and *Nanos2*$^{TAG/TAG}$ SSC lines, were used for the experiments.

SLAM-seq libraries were prepared by following the manifacturers' instructions in the QuantSeq 3′ mRNA-Seq Library Prep Kit FWD for Illumina, SKU: 015.24, Lexogen.

2.16.1 SLAM-seq analysis

The SLAM-seq bioinformatic analysis was performed by Dr. Louie N. van de Lagemaat, as described in the SlamDunk pipeline (Herzog et al., 2017). T to C conversion rates obtained from different timepoints were normalised to 'Time 0' for each transcript. These values were used to fit a first order decay reaction with the R stats package.

2.16.2 Sylamer analysis

Sylamer analysis was performed as described in (van Dongen et al., 2008). As input for the analysis, we used the 3' UTRs of mRNAs of known stability (calculated with SLAM-seq), ranked from short to long half-life. 3' UTRs were selected by choosing the average length of all possible 3' UTRs for a given transcript.

2.17 DNA quality controls

The quality of DNA libraries was assessed with the High Sensitivity D1000 ScreenTape (5067-5584, Agilent Technologies). DNA samples concentration was measured with Qubit dsDNA HS Assay Kit (Q32854, Life Technologies).

2.18 Sequencing

Libraries generated by RNA and SLAM-seq were sequenced with an Illumina HiSeq2500 on a 50 bp, Single-end run.

Libraries generated by CRAC were sequenced with an Illumina NextSeq500, high output (400 million reads), on a 75 bp, Single-end run. All sample were sequenced at EMBL Heidelberg GeneCore Facility.

Chapter 3. Identification of NANOS2 occupancy on RNA in SSCs

3.1 Introduction

Nanos family genes exert essential functions during development. These vary from body pattern formation in *Drosophila*, to their evolutionary conserved roles in germ cells (De Keuckelaere et al., 2018). A common feature among NANOS proteins is their CCHC zinc finger motifs, located in the C-terminus. Structural studies have demonstrated that the zinc fingers mediate protein-RNA interactions (Hashimoto et al., 2010b; Weidmann et al., 2016). Importantly, the zinc fingers are essential to support all known NANOS functions, in *Drosophila* (Arrizabalaga and Lehmann, 1999). It is generally assumed that also *Nanos2* functions in mouse strictly depend on its ability to bind to RNA. Therefore, different attempts have been made to characterise NANOS2 targets, both in male gonads and post-natal SSCs, by using native RIP (RNA Immunoprecipitation) (Saba et al., 2014; Suzuki et al., 2010; Zhou et al., 2015). However, native RIP has several technical limitations (see Chapter 1), which did not allow to unequivocally assess the following aspects, regarding protein-RNA interaction: first, whether NANOS2 contacts directly RNA, or through other bridging proteins; second, the precise identity of NANOS2 targets at the transcriptomic level; lastly, whether NANOS2 binds to RNA with sequence specificity.

Therefore, I sought to address the issues above, and to provide a rigorous and extensive characterisation of NANOS2-RNA interactions. To do so, I employed one of the most stringent method, called CRAC (Granneman et al., 2009b) (see Chapter 1), and I applied it to *in vitro* SSC lines. In this chapter, I describe how we generated a *Nanos2*^{TAG} mouse model for CRAC. Then, I assessed that the presence of the TAG at NANOS2 N-terminus did not alter protein localisation and function, in mouse SSCs. I optimised CRAC in HEK293t cells

and then performed it in SSC lines. In this chapter, I also show extensive CRAC analysis, including *de novo* motif discovery, which provides unprecedented insight into NANOS2 transcriptomic occupancy in SSCs. Eventually, I illustrate the biological identity of NANOS2 RNA targets, by performing gene ontology analysis, and by using Markov clustering algorithm.

3.2 Generation of the *Nanos2^{TAG}* mouse allele for CRAC

Nanos2 locus is composed of a single exon and spans 1.563 Kb on chromosome 7. The locus consists of a 5' UTR, followed by a short coding sequence (411 bases), and a 3' UTR (Figure 3.1 A). Therefore, introns are not present in this gene.

In order to perform CRAC, we engineered a *Nanos2^{TAG}* mouse allele by gene targeting in mouse embryonic stem cells (ESCs). A Tag cassette was inserted into the endogenous *Nanos2* locus, after the starting codon of *Nanos2*, to produce an N-terminal fusion protein (TAG-NANOS2). We designed a variant of the original CRAC Tag, to allow both complex biochemistry and also *in vivo* localisation studies. Thus, the Tag consisted of five moieties: V5-cMyc-PreScission-6xHis-eGFP. The first four components comprised a set of epitopes to enable multiple biochemical purifications of NANOS2-RNA complexes, by CRAC. These included the V5 and c-Myc epitope tags, useful to perform endogenous NANOS2 immunoprecipitation; the PreScission cleavage site, which was used for protein cleavage; and the stretch of polyhistidine residues (6xHis), used for protein purification with nickel beads. Following these tags, an eGFP was also inserted to assess the *in vivo* localisation of the protein, by FACS and immunofluorescence.

First of all, the Tag was inserted into a targeting vector (*pDTA-Tag-Frt-Neo-Frt*), which carried homology arms to the Nanos2 locus, to allow precise insertion of the cassette into the desired genomic location. After the *Tag*, the vector contained a neomycin (*neo*) resistance gene (white rectangular), which was used to select recombinant ESC clones in culture. Additionally, the *neo*

cassette was flanked on both sides by flippase recognition target sites (FRT) (green triangles), to enable subsequent excision of the *neo* cassette, by a flippase (FLP).

The targeting vector described above was electroporated into A9 ESCs, which were derived from a F1 hybrid, with mixed Bl6N and 129 genetic background. Homologous recombinant clones were enriched by addition of the G418 antibiotic in the culture media. Surviving clones were picked and screened for the presence of the transgene by Southern blot, which was performed by Dr. Ivanova. Genomic DNA was extracted and digested with a NheI restriction enzyme. The *Nanos2*^{WT} locus contains two NheI restriction sites, spanning a region of 6.6 Kb. However, the recombinant *Nanos2*$^{TAG\text{-}neo}$ locus, contains three NheI sites, since the integration of the *Tag-neo* cassette introduced an additional one (in brown) after the last Frt sequence. The sequence enclosed between the last two NheI sites in *Nanos2*$^{TAG\text{-}neo}$ spans 3.8 Kb. Therefore, different DNA digestion patterns are expected from *Nanos2*^{WT} and *Nanos2*$^{TAG\text{-}neo}$ clones. Southern blot revealed the presence of two clones with two bands of 6.6 and 3.8 Kb, representing the *Nanos2*^{WT} and the *Nanos2*$^{TAG\text{-}neo/+}$ alleles, respectively (Figure 3.1 B).

After confirming the incorporation of the *Tag-neo* cassette by Southern Blot, *Nanos2*$^{TAG\text{-}neo/+}$ clones were expanded and injected into embryos, to generate chimeric mice. These were crossed with BL6N mice to generate founder mice, which were examined for the presence of the transgene. Founder mice were crossed with BL6N to assess the germ-line transmission of the allele to the F1 generation. Afterwards, *Nanos2*$^{TAG\text{-}neo/+}$ F1 mice were crossed with mice expressing Flp (FLPeR, (Farley et al., 2000)), to induce the FLP-FRT recombination and thus the excision of the *neo* cassette. These matings resulted in the generation of the *Nanos2*$^{TAG/+}$ mouse allele.

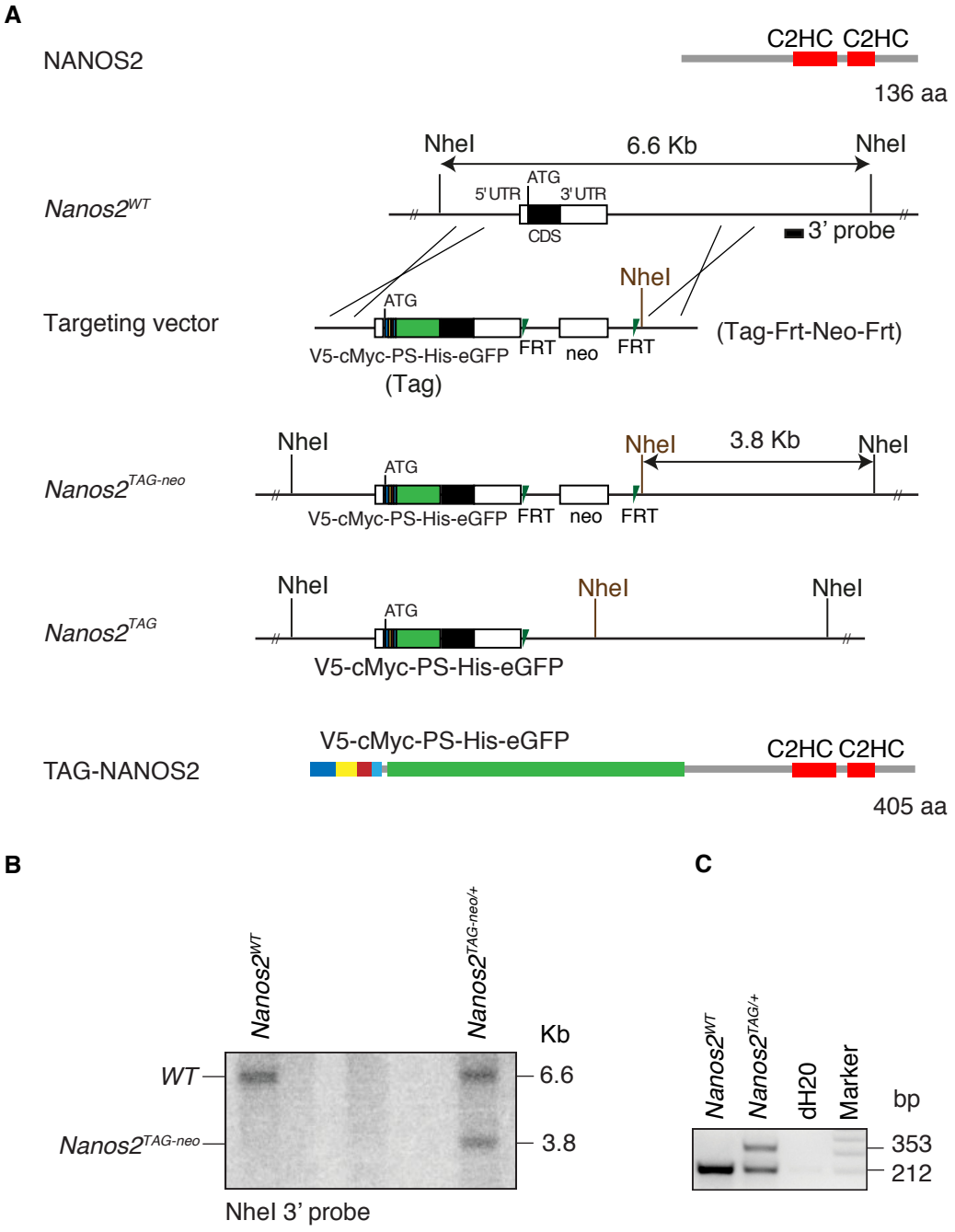


Figure 3.1 Targeting strategy used for the generation of the Nanos2TAG mouse allele.

A. Schematic representation of the Wild-type NANOS2 protein, with the two zinc finger domains (CCHC) in red. Amino acid sequence length (aa) is also shown. The *Nanos2*^{WT} locus is composed of a protein-coding region (CDS) (black box) and a 5' and a 3' UTRs (white boxes). In the targeting vector the multi-coloured boxes represent the Tag, which is composed of V5-cMyc-PS-His-eGFP (PS is PreScission).

The white box *neo* is the is the Neomycin gene and green triangles represent the FRT sites. The resulting N-terminal TAG-NANOS2 fusion protein is also showed and represented in scale. Wild-type NheI restriction sites are in black, whereas the NheI site introduced with the targeting vector is in brown. Also, the 3' probe used for Southern blot is shown as a black box. **B.** Example of Southern blot performed on genomic DNA from two ESCs clones. **C.** Nanos2 PCR-genotyping using biopsies from *Nanos2*^{WT} and *Nanos2*$^{TAG/+}$ mice. The molecular size of the PCR products is expressed in base pair (bp).

3.3 Functional validation of the *Nanos2*^{TAG} mouse line

Fusion proteins are useful tools to investigate proteins function. However, they can also negatively affect cells, since they have a modified amino acid sequence (and probably folding) compared to their wild-type counterparts.
I have previously shown how we engineered a *Nanos2*^{TAG} allele (Figure 3.1), which is supposed to generate an N-terminal fusion protein, called TAG-NANOS2. Since this protein is three times bigger than its wild-type counterpart (~47 kDa and ~14 kDa, respectively), we verified whether TAG-NANOS2 was functional. NANOS2 is required for the maintenance of male gonocytes and also it is essential to sustain male fertility over time (see Chapter 1); therefore, I assessed whether TAG-NANOS2 could also do so. To this end, I evaluated male fertility in detail by assessing whether TAG-NANOS2 could support all stages of spermatogenesis. To this end, I collected testes from *Nanos2*$^{TAG/+}$ and *Nanos2*$^{TAG/TAG}$ mice, measured their weight, and performed histological analysis of testes cross-sections. Since *Nanos2*$^{-/-}$ mice lose germ cells progressively over time (Tsuda et al., 2003), I also analysed testes from six months old mice, to detect any detrimental effect in the long-term. Importantly, *Nanos2*$^{TAG/TAG}$ mice had normal testicular weight, compared to the control *Nanos2*$^{TAG/+}$ (Figure 3.2 B). Moreover, the histological evaluation of testicular cross-sections showed that *Nanos2*$^{TAG/TAG}$ mice also had a healthy tubular

morphology (Figure 3.2 C). Indeed, I could not detect empty tubules and importantly, I could observe all spermatogenic stages within the seminiferous tubules. Therefore, TAG-NANOS2 protein was able to sustain spermatogenesis fully, also in the long term.

Next, we assessed whether *Nanos2*$^{TAG/TAG}$ mice had normal litter size. Thus, we paired adult homozygous *Nanos2*$^{TAG/TAG}$ males with wild-type females for three months and counted the number of litters they generated. Control mice were represented by *Nanos2*$^{TAG/+}$ males, which were bred with wild-type females. In the graph shown in Figure 3.2 D, each black dot represents the average litter size per adult male, whereas red lines are the variance, calculated with a T-test. Importantly, homozygous males displayed standard litter size, when compared to controls.

Therefore, these data indicated that adult *Nanos2*^{TAG} males were completely fertile.

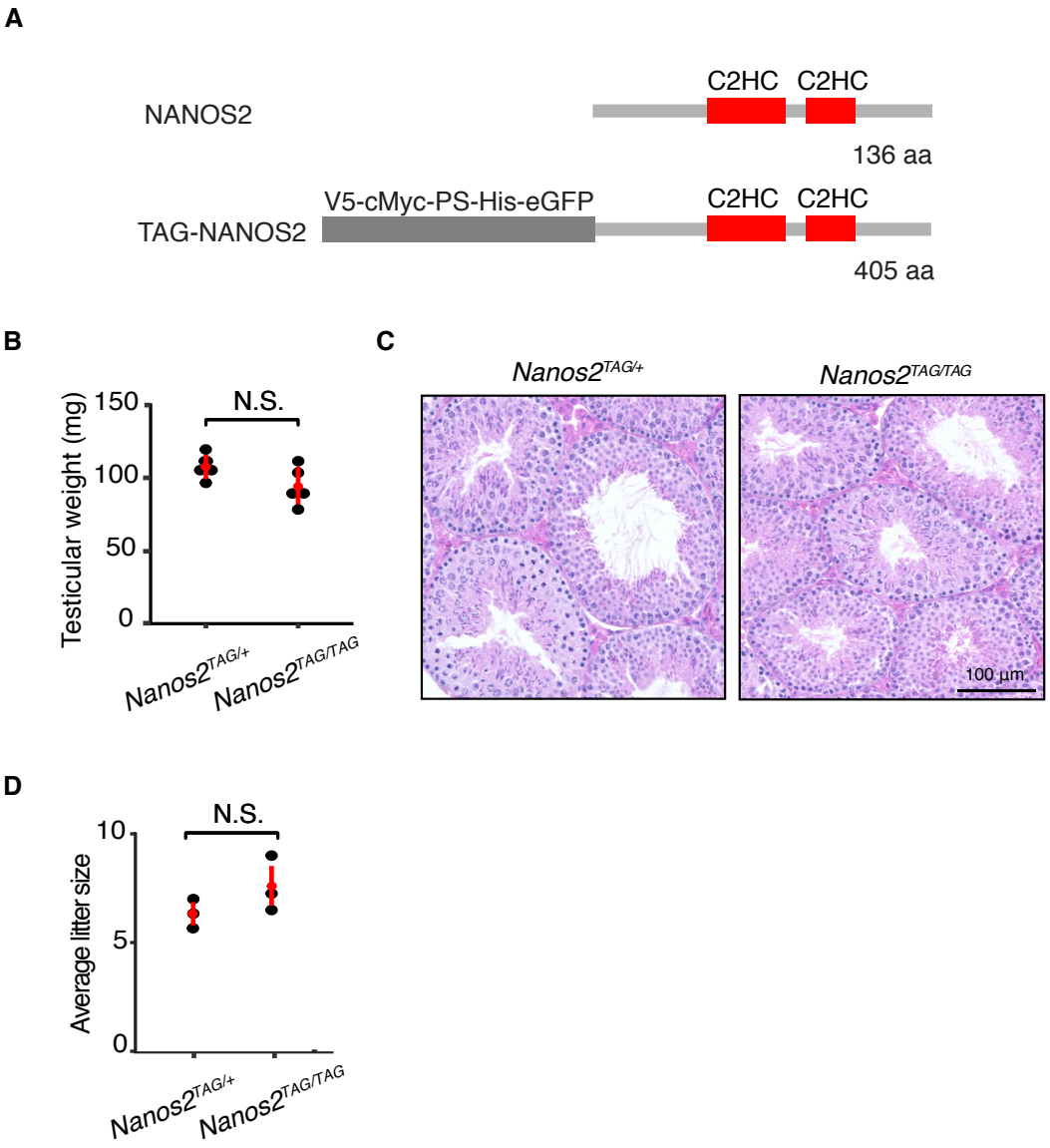


Figure 3.2 Assessment of fertility and spermatogenesis of the *Nanos2*^{TAG} mouse allele.

A. Schematic representation of the wild type NANOS2 and TAG-NANOS2 proteins. Lengths of amino acid sequences are also shown (aa). **B.** Dot plot showing the testicular weight (in mg) of *Nanos2*$^{TAG/+}$ and *Nanos2*$^{TAG/TAG}$ mice, where *Nanos2*$^{TAG/+}$ represents the control. Mice were six months old. **C.** Representative images of testes cross-sections, from six months old mice. Testes are stained with Hematoxylin and Eosin and scale bar is 100 μm. **D.** Dot plot showing the average litter size of *Nanos2*$^{TAG/+}$ and *Nanos2*$^{TAG/TAG}$ mice. In **B.** and **D.,** mean and s.d. are shown in red.

Two tailed Student's t-tests were used to compare differences between *Nanos2*$^{TAG/+}$ and *Nanos2*$^{TAG/TAG}$, assuming equal variance.

Afterwards, I assessed whether the presence of the TAG affected NANOS2 localisation *in vivo*. To this end, I performed whole-mount immunofluorescence of seminiferous tubules, isolated from adult *Nanos2*$^{TAG/TAG}$ mice. Images of these samples were acquired with a confocal microscope and further analysed.

First of all, I examined the intracellular localisation of TAG-NANOS2, which is known to be cytoplasmic. To detect TAG-NANOS2 expression, I stained tubules with an Anti-GFP antibody, and to visualise nuclei, I also stained them with Dapi. As shown in Figure 3.3 A, TAG-NANOS2 (GFP signal in green), is expressed only in the cytoplasm, and no signal is found in the nucleus. Thus, the presence of the TAG did not disturb NANOS2 intracellular localisation, *in vivo.*

Moreover, I examined TAG-NANOS2 localisation within the SSC compartment, which is constituted by a tiny subset of the undifferentiated spermatogonia (see Chapter 1). Thus, I analysed TAG-NANOS2 expression in relation to GFRα1, a specific marker for self-renewing SSCs (see Chapter 1), by using Anti-GFP and Anti-GFRα1 antibodies (Figure 3.3 B). Importantly, the majority of NANOS2-positive SSCs (in green) also expressed GFRα1 (in red), and they were mainly represented by single (A_{single}, A_s) and short chains of cells, such as A_{paired} (A_{pr}) and $A_{aligned4}$ (A_{al4}). Importantly, these cells are thought to contain most of the self-renewing SSCs pool (see Chapter 1). These observations were also confirmed by quantifying the number of SSCs which expressed GFP (Gfra1^{Neg}; Nanos2^{Pos}), GFRα1 (Gfra1^{Pos}; Nanos2^{Neg}) or both of them (Gfra1^{Pos}; Nanos2^{Pos}) (Figure 3.3 C). These cells were also classified according to their topology, as A_s, A_{pr}, A_{al4}, A_{al8}, A_{al12}. I also normalised SSC counts by the number of Sertoli cells, which were found in the same portions of tubule. Sertoli cells are the only somatic cell type present within

seminiferous tubules (see Chapter 1), and their number is constant along tubules' length. Sertoli cells were identified by analysing nuclei morphology, stained by DAPI, as these display two characteristic heterochromatic masses near the nucleolus. Overall, SSCs quantification confirmed our previous observation that most of the Nanos2^{Pos} cells overlapped with the GFRα1 pool. Importantly, these data are in accordance with previously published results about NANOS2 localisation *in vivo* (Suzuki et al., 2009).

Moreover, I assessed whether TAG-NANOS2 was correctly confined to the undifferentiated spermatogonia compartment. Hence, I co-stained tubules with an Anti-GFP (for TAG-NANOS2) and an Anti-c-Kit antibody, which detected c-Kit, a marker of differentiating spermatogonia. Importantly, as shown by Figure 3.3 D, TAG-NANOS2 (in green) was never expressed in c-KIT-positive cells (in red), which were only represented by long chains of spermatogonia. This indicated that the TAG-NANOS2 fusion protein was not expressed in the differentiated spermatogonia.

In summary, these analyses indicated that the presence of the TAG in the TAG-NANOS2 fusion protein did not impact fertility nor spermatogenesis, even in the long term. Importantly, NANOS2 protein localisation *in vivo* was not altered by the TAG, since TAG-NANOS2 maintained its typical cytoplasmic expression and it was correctly confined to the SSC compartment.

Therefore, these data demonstrated that the *Nanos2*$^{TAG/TAG}$ mouse allele was functional, and that could be used for CRAC.

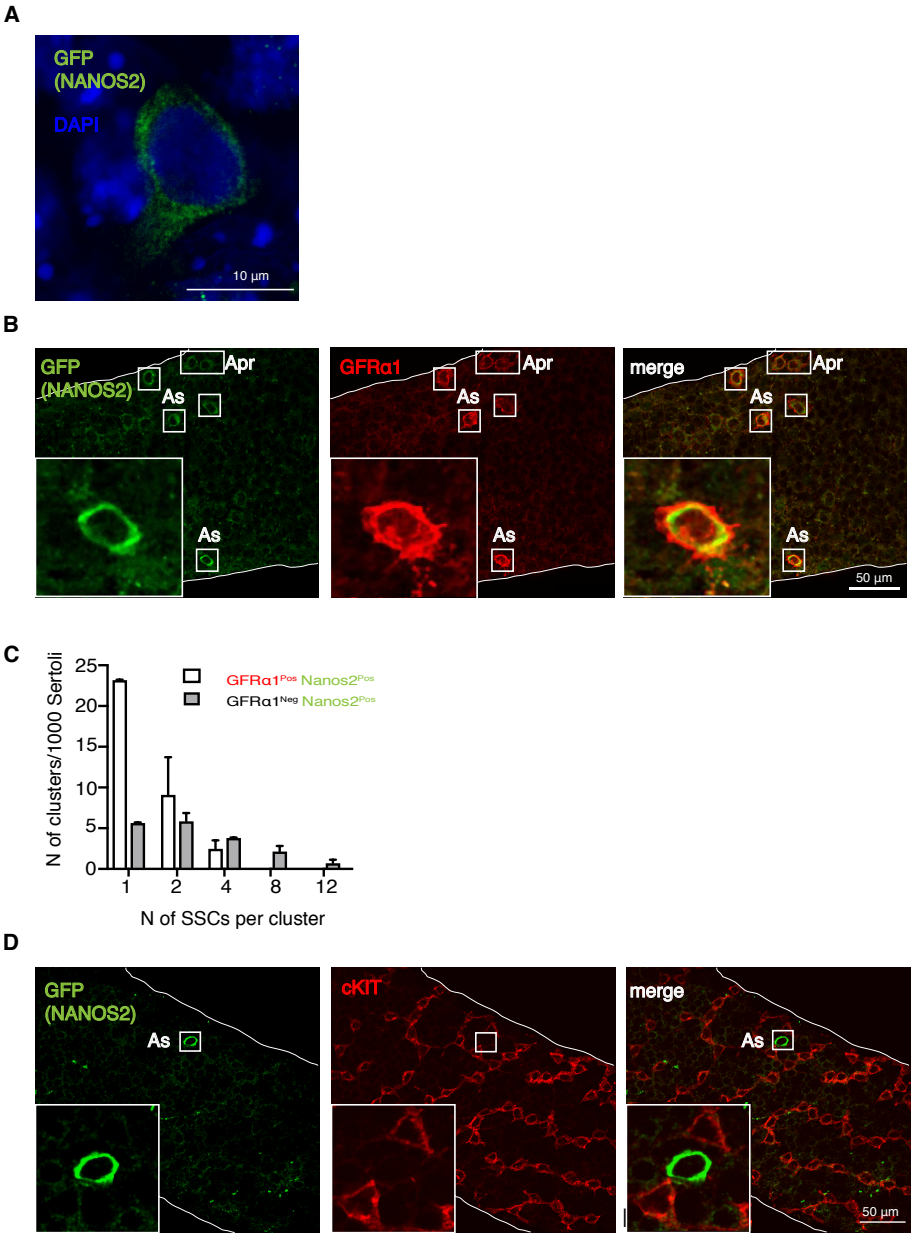


Figure 3.3. Assessment of NANOS2 localisation by whole-mount immunofluorescence (IF) of seminiferous tubules, from *Nanos2*$^{TAG/TAG}$ mice.

A. Representative IF image showing NANOS2 (Anti-GFP) intra-cellular localisation. Nuclei are stained with DAPI. **B.** Representative IF images of NANOS2 and GFRα1. As and Apr (Apaired) SSCs are shown. **C.** Bar graph showing number of GFP positive and GFRα1 positive SSCs per cluster (1=A_s, 2=A_{pr}, 4=Aal_4, 8=Al_8, 12=Al_{12}). SSCs counts were normalised for the number of Sertoli cells (y-axis). **D.** Representative IF

images showing NANOS2 and c-KIT localisation in spermatogonia. NANOS2 is expressed in an As (Asingle) SSC. All images have been acquired with a confocal microscope.

3.4 SSCs derivation for CRAC

It is estimated that SSCs constitute only 0.03% of all germ cells, in the adult mouse testis (Tagelenbosch and de Rooij, 1993). Furthermore, only a tiny subset of SSCs express Nanos2, approximately three or four thousand cells per animal. However, CLIP methods, like CRAC, require at least 10^8-10^9 cells as starting material (Ramanathan et al., 2019). Therefore, the rarity of Nanos2 positive SSCs in testis hinders their molecular and biochemical characterisation in *ex vivo* isolated cells, by CRAC. To circumvent this problem, I derived SSCs from neonatal mouse testis and expanded them *in vitro*, as initially described in 2003 (Kanatsu-Shinohara et al., 2003). SSC lines were derived from *Nanos2*^{WT}, *Nanos2*$^{TAG/+}$ and *Nanos2*$^{TAG/TAG}$ pups, at postnatal day 7. Since *Nanos2*^{WT} lines were used as control for the all rest of the experiments, from now onwards, they will be called *Nanos2*^{CTL}, whereas *Nanos2*$^{TAG/TAG}$ lines will be named *Nanos2*^{TAG}.

As shown in Figure 3.4 A, SSCs in culture grow as grape-shaped colonies, on MEF feeders, and they can be amplified almost indefinitely *in vitro*, without losing their self-renewal ability (see Chapter 1 and 2).

To control NANOS2 expression in these cells, I prepared protein extracts from *Nanos2*^{CTL} and *Nanos2*^{TAG} lines and performed western blotting. By taking advantage of the GFP inserted into the TAG, I probed the membrane with an Anti-GFP antibody, to detect TAG-NANOS2. A band of approximately 47 kDa, corresponding to the expected TAG-NANOS2 molecular size, was observed in *Nanos2*^{TAG} sample (Figure 3.4 B). Importantly, no GFP signal was detected in the control (*Nanos2*^{CTL}), although both samples displayed comparable levels of Actin, which indicated that a similar amount of protein extracts were used.

Afterwards, I analysed TAG-NANOS2 expression at single-cell level, by looking at the GFP signal by flow cytometry. SSCs were purified from MEFs and selected according to size and shape. Only single and living cells were used for the analysis. Importantly, *Nanos2*^{TAG} displayed a mono-modal GFP distribution (Figure 3.4 C), and an average GFP intensity higher than *Nanos2*^{CTL}. However, a partial overlap was observed between *Nanos2*^{CTL} and *Nanos2*^{TAG} populations. This indicated that a small subset of *Nanos2*^{TAG} SSCs did not express GFP (TAG-NANOS2), since its level was comparable to the autofluorescent signal in the negative control.

In summary, TAG-NANOS2 was successfully expressed in the *Nanos2*^{TAG} line and had the expected molecular size. Moreover, flow cytometry analysis confirmed the expression of TAG-NANOS2 at the single cell level, in the great majority of SSCs. Only a small subset of *Nanos2*^{TAG} cells did not significantly express GFP. This could be explained by a certain degree of heterogeneity associated with SSC lines.

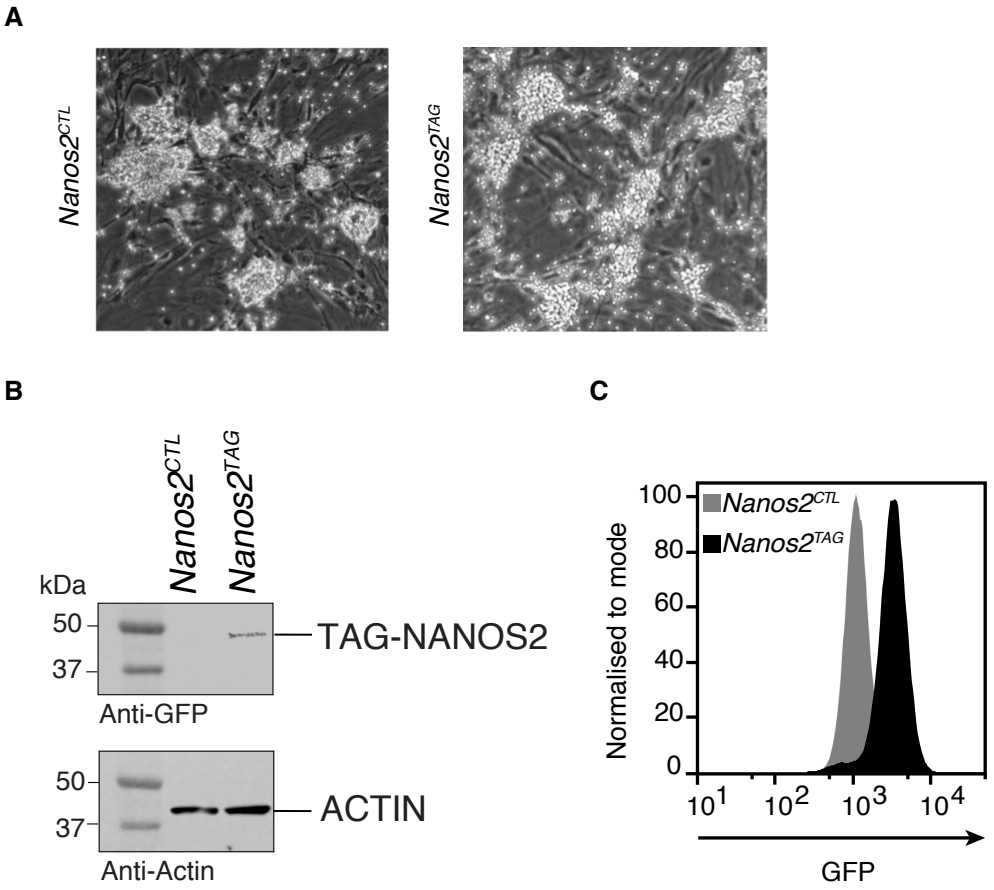


Figure 3.4 Derivation and characterisation of *in vitro* SSC lines.

A. Representative bright field images of SSC lines derived from *Nanos2*^{CTL} (*Nanos2*$^{+/+}$) and *Nanos2*^{TAG} (*Nanos2*$^{TAG/TAG}$) juvenile testes. SSCs are the bright, grape-shaped colonies, grown on top of MEF feeders. **B.** Western-blot showing endogenous NANOS2 expression (GFP signal) in *Nanos2*^{TAG} SSCs. Anti-Actin signal is also shown. **C.** Histogram from flow cytometry analysis, showing the normalised GFP intensity for *Nanos2*^{CTL} and *Nanos2*^{TAG} populations.

3.5 NANOS2 CRAC optimisation in HEK293t cells

When performing CRAC, continuous supply of biological material might be required at the beginning, in order to identify the best experimental conditions for the RBP of interest.

However, SSCs are slow-cycling cells (division rate of seven days), and they express endogenous levels of NANOS2. Therefore, in order to obtain sufficient amounts of cells in a reasonable amount of time, I decided first to optimise CRAC in HEK293t cells and, only afterwards, to apply it to SSC lines.

To this end, I over-expressed TAG-NANOS2 in HEK293t cells (*Nanos2*^{OE}), by cloning the Tag-eGFP-Nanos2 construct into a PyCAG vector (Chambers et al., 2003), between the NotI and XhoI restriction sites (Figure 3.5 A). Three 150 cm^2 plates were transiently transfected with the Tag-eGFP-Nanos2-PyCAG vector, and the expression of TAG-NANOS2 protein was examined under a fluorescent microscope. As it is shown in Figure 3.5 B, upon transfection, HEK293t could successfully express TAG-NANOS2 (GFP signal, *Nanos2*^{OE}). Afterwards, I tried to understand whether the TAG-NANOS2 fusion protein was suitable for CRAC, by testing the efficiency of the main protein purification steps: immunoprecipitation, cleavage, and affinity binding to Nickel beads (see Chapter 1). Therefore, I collected WT and *Nanos2*^{OE} cells forty-eight hours after transfection and prepared protein extracts. These were incubated with an Anti-V5 antibody, to immunoprecipitate TAG-NANOS2. Then, to remove the Anti-V5 antibody together with the V5 tag, I added the 3C Protease to the beads. After cleavage, I incubated samples with the Nickel-agarose resin (Ni-NTA), performed the washes, and eluted proteins. Samples were subjected to SDS-PAGE and western blot, whose results are shown in Figure 3.5 C. In the Input (first lane), TAG-NANOS2 was expressed at the expected molecular size, ~47 kDa, and it was efficiently immunoprecipitated since no protein was left in the flow through (second lane, FT). Importantly, the immunopurified TAG-NANOS2 was successfully cleaved, since the removal of both V5 and PreScission tags reduced its molecular size to ~43 kDa (third

lane, Cleaved). Additionally, only minimal fractions of both full length and cleaved proteins were left on the beads (fourth lane). Moreover, cleaved TAG-NANOS2 could efficiently bind the Ni-NTA resin, as no protein was detected in the flow through (FT) and the great majority was found in the eluted sample (second last lane, Cleaved). Only a minimal part of the sample was left on the beads (last lane).

Overall, these data showed that TAG-NANOS2 could successfully undergo the different purification steps required for CRAC, under the conditions tested above.

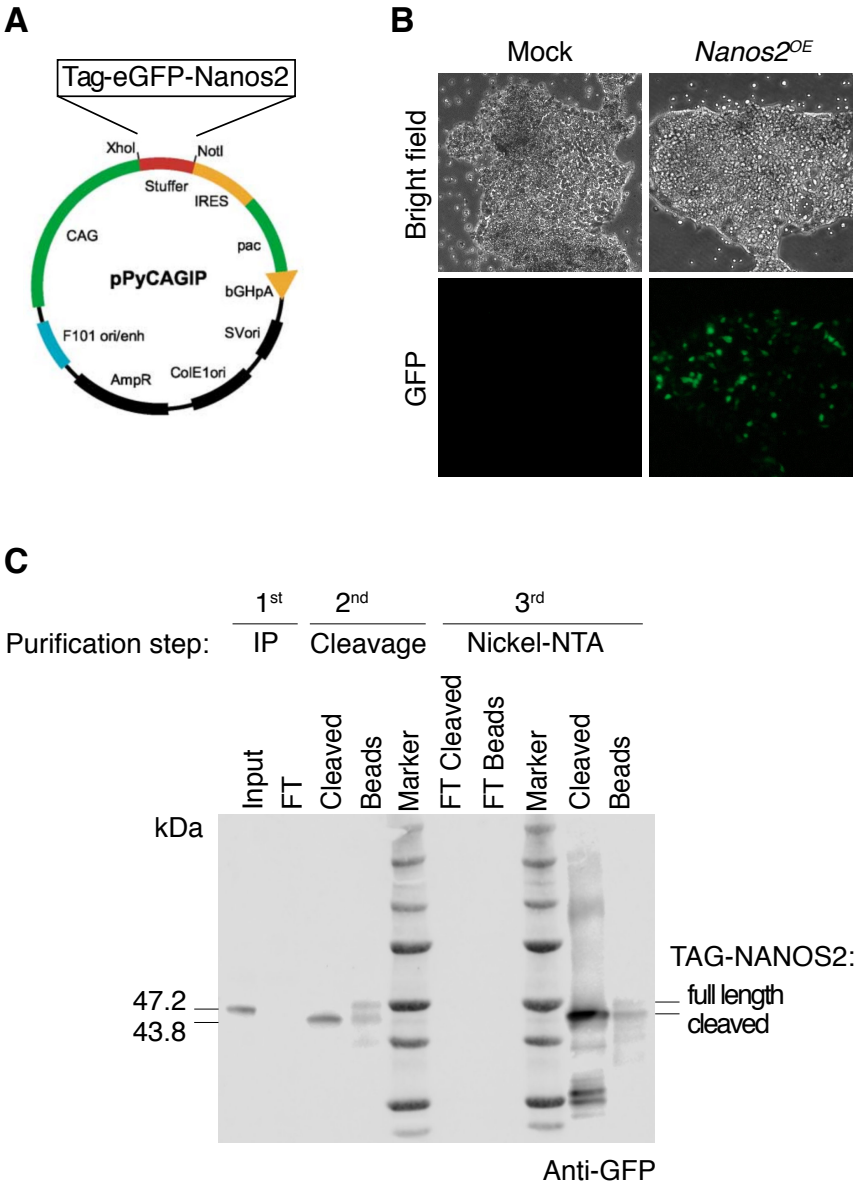


Figure 3.5. Generation of *Nanos2^{OE}* HEK293t and CRAC optimisation.

A. Cloning strategy used to generate *Nanos2^{OE}* HEK293t. The 'Tag-eGFP-Nanos2' construct was inserted into the pyCAG vector, between the XhoI and the NotI restriction sites. **B.** Transfection of the 'Tag-eGFP-Nanos2' vector into HEK293t cells. Mock is the control. *Nanos2^{OE}* cells show GFP signal (in green), which indicates transfected cells. **C.** Western blot showing NANOS2 immunoprecipitation (IP), cleavage and binding to Nickel-NTA beads. 'FT' indicates flow through. 'Marker' is the protein marker used.

Next, I performed the complete CRAC protocol with HEK293t cells. After transient transfection, I cross-linked control (Mock) and *Nanos2*^{OE} cells with 254-nm UV light, at 400 mJ/cm^2, a medium-high dose. After protein immunoprecipitation and cleavage, I trimmed RNA by partial digestion with RNase A and T1. I coupled protein-RNA complexes to Nickel-NTA beads and immobilised them on spin columns. I applied denaturing washes, by using a buffer containing GuHCl. While still on the columns, RNA was subjected to phosphatase treatment, to remove phosphate groups previously left by the RNase. RNA was radiolabelled and 3' RNA adapters were ligated to RNA, to enable library generation afterwards. Each sample was also barcoded with a unique 5' adapter, to allow their univocal identification afterwards. After eluting the RNA-protein complexes from the columns by using an excess of Imidazole, I performed SDS-PAGE and transferred samples to a nitrocellulose membrane. The radioactive membrane was exposed for autoradiography, and the film analysed following 1 and 24 hours (Figure 3.6 A). Strikingly, the autoradiogram showed a very strong signal corresponding to TAG-NANOS2, already after 1 hour. Moreover, this signal was specifically associated with *Nanos2*^{OE} sample, as the control did not show any band, even after 24 hours. Importantly, the signal was extended above NANOS2 molecular size (up to ~75 kDa), showing that RNA was successfully trimmed by RNase. Furthermore, I did not detect other strong bands at different sizes, suggesting that I did not co-purify other RBPs together with TAG-NANOS2. This also suggested that the UV-cross-linking dose used was not too high. Therefore, these data indicated that NANOS2 had a strong and specific RNA-binding ability in HEK293t cells.

Next, I extracted the RBP-RNA complexes from the membrane, by selecting only the area which had a radioactive signal on the film. To remove proteins, I treated samples with Proteinase K, and then I extracted and reverse transcribed RNA. To generate libraries, I amplified the cDNA with PCR primers complementary to the 5' and 3' adapters previously used. PCR were performed

from the *Nanos2*^{OE} sample and from the untagged HEK293t line (WT), which was the CRAC negative control. Additionally, to verify that the PCR mixture had no contaminants, another control was included, by performing a PCR with water instead of the cDNA template (dH20). PCRs were performed by using only 22 cycles and the PCR products controlled in a high resolution 3% MetaPhor agarose gel. As shown by Figure 3.6 B, libraries were successfully amplified only from the *Nanos2*^{OE} sample and not from the negative controls. Indeed, I could detect a DNA smear (in grey, dotted red rectangular) which had the expected library size (between ~130 and ~190 bp). Below the smear, it is also evident an intense black band (arrowhead) represented by adapter-dimers (~130 bp), which did not contain an RNA insert. These were excluded by performing gel extraction. After quantification of the purified libraries, I could conclude that TAG-NANOS2 could indeed precipitate significant amount of RNA, by CRAC.

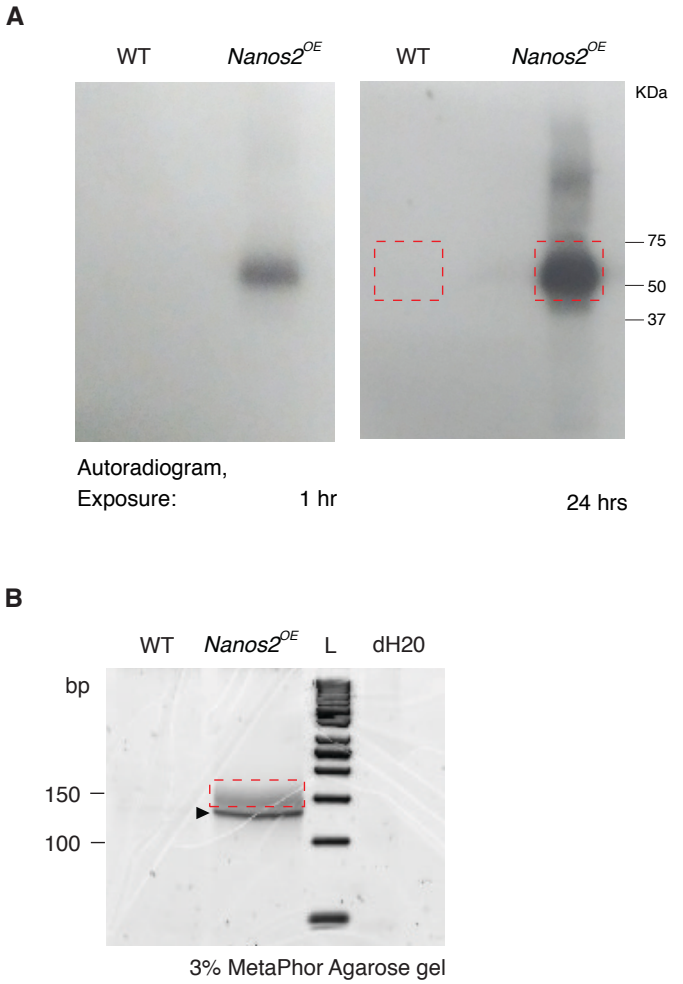


Figure 3.6. NANOS2 CRAC in HEK293t cells.

A. Autoradiogram showing the radioactive signal associated to NANOS2-RNA complexes in *Nanos2*^{OE} cells. Two films are shown. The one on the left was exposed for 1 hour and the other one for 24 hours (right). Red dotted rectangles indicate the area from which protein-RNA complexes were extracted. **B.** cDNA libraries amplified by PCR, using 22 cycles, were run in a 3% MetaPhor Agarose gel. Red dotted rectangle indicates the area excised from the gel, to purify libraries from adapter-dimers (arrowhead). 'WT' sample represents the untagged Nanos2 HEK293t (CRAC negative control), and 'dH20' sample represents a PCR amplified with water instead of the cDNA template (PCR negative control). 'L' is DNA ladder.

3.6 NANOS2 CRAC in SSC lines

After optimising NANOS2 CRAC in HEK293t cells, I proceeded with CRAC in *in vitro* SSC lines. First of all, I tested the protein immunoprecipitation from a *Nanos2*$^{TAG/TAG}$ 10 cm^2 plate, to estimate the amount of material that I would need for the experiment.

After preparing protein extracts and performing SDS-PAGE, I ran a western blot and probed the membrane with an Anti-V5 antibody. As shown in Figure 3.7 A, NANOS2 could be immunoprecipitated specifically from *Nanos2*$^{TAG/TAG}$ SSCs and not from the control. However, the efficiency of the IP was not high, since a fraction of NANOS2 was still found in the flow through (FT). Subsequently, *Nanos2*^{CTL} and *Nanos2*$^{TAG/TAG}$ lines were grown in eight 150 cm^2 plates, and CRAC was performed with the same conditions used for HEK293t. Results from the autoradiography are shown in Figure 3.7 B. Strikingly, a faint band with the molecular size of TAG-NANOS2 was already visible on the film after a few hours, but became evident after 24 hours. Additionally, the smear above TAG-NANOS2 indicated a partial RNA digestion. Importantly, almost no background signal was observed in the *Nanos2*^{CTL}, indicating that NANOS2 specifically bound to RNA, in mouse SSC lines. The absence of strong bands on the film, at other molecular sizes, suggested that TAG-NANOS2 did not co-purify other RBPs, as expected after using very stringent purification conditions.

Therefore, this data indicated that NANOS2 bound directly to RNA in SSCs.

To prepare libraries, RNA was purified and reverse transcribed. First of all, I tried to amplify the cDNA by PCR with 22 cycles, as previously done for the CRAC in HEK293t. However, after running the PCR product in a MetaPhor agarose gel, I could detect only a weak DNA signal (Figure 3.7 C, black dotted rectangle), although this present only in the *Nanos2*^{TAG} sample. Additionally, after gel-purifying this library, I did not obtain enough material for sequencing. Thus, I repeated the PCR using 28 cycles, and the same was done for the negative control. This time I could detect a clear DNA smear, which was

present only in the *Nanos2*^{TAG} sample (Figure 3.7 D). This result indicated that although I greatly increased the number of PCR cycles, I did not amplify non-specific DNA from the negative control. These libraries were gel-purified in order to exclude adapter-dimers and sent for sequencing.

Afterwards, I repeated the experiment another time from the same cell lines (technical replicate) and importantly, I could confirm the same autoradiogram and PCR results. Moreover, I also performed CRAC from other two SSC lines, *Nanos2*^{CTL} and *Nanos2*$^{TAG/+}$ (biological replicate). Strikingly, also this experiment validated previous results, reinforcing the fact that NANOS2 binds directly and specifically to RNA in SSCs.

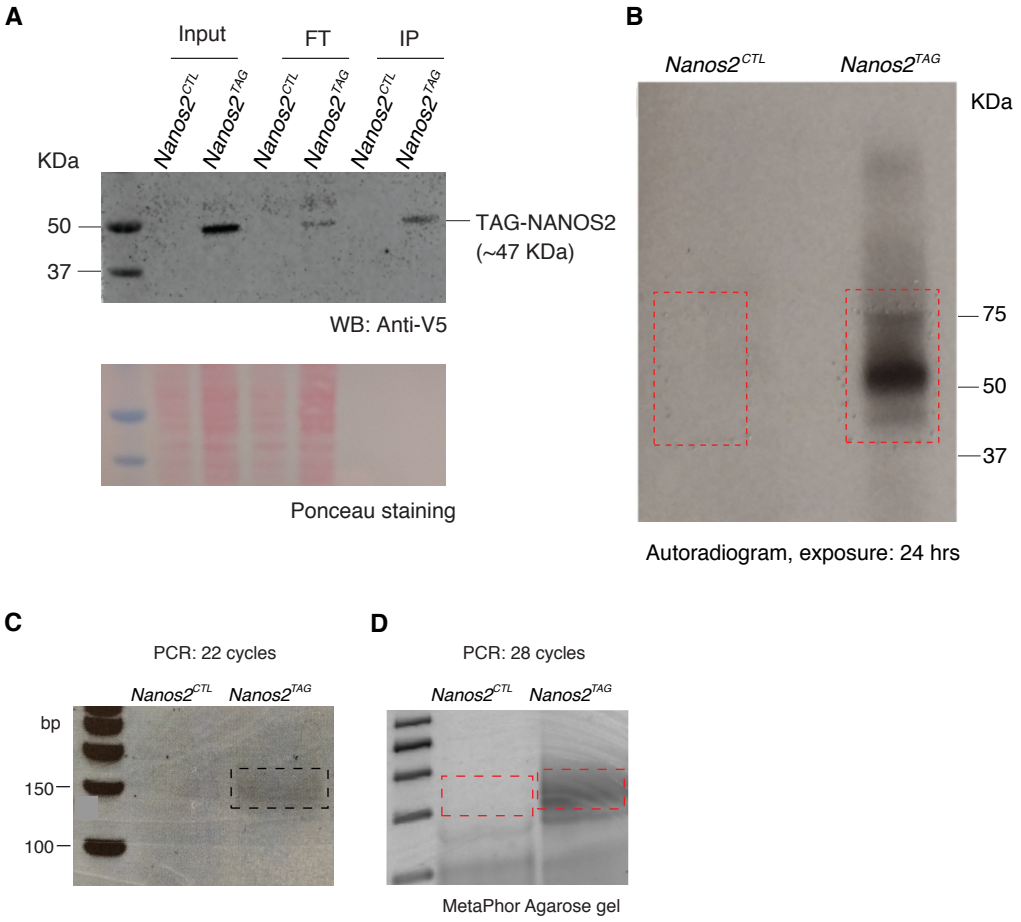


Figure 3.7. NANOS2 CRAC in SSC lines.

A. Endogenous NANOS2 immunoprecipitation (IP) from *Nanos2*^{CTL} (*Nanos2*$^{+/+}$) and *Nanos2*^{TAG} (*Nanos2*$^{TAG/TAG}$) SSC lines. Input, flow through (FT) and IP samples are shown in the blot. The same membrane was also stained with Ponceau as a quality control for IP-western-blot. **B.** Autoradiogram showing the radioactive signal associated to NANOS2-RNA complexes in *Nanos2*^{TAG}. Red dotted rectangles indicate the area from which protein-RNA complexes were extracted and purified from the membrane. The film was exposed for 24 hours. **C.** PCR optimisation to amplify CRAC libraries from cDNA. The PCR product amplified with 22 cycles was analysed in an agarose gel. Black dotted rectangle indicates the library (faint smear) which was gel-purified. **D.** cDNA was amplified again by using 28 cycles of PCR. Red dotted rectangles indicate the libraries extracted from the gel and used for sequencing afterwards.

3.7 NANOS2 CRAC analysis

3.7.1 Reads filtering and quality controls

Four CRAC libraries were subjected to high-throughput sequencing using an Illumina NextSeq500, on a single-end run. Libraries included two independent replicates (technical replicates) from a *Nanos2*^{TAG} SSC line, as well as two negative controls from a *Nanos2*^{CTL} line. High-depth sequencing generated a total of 427 million reads, ranging from 0.76 million and 209.8 million reads, depending on the sample (Table 3.1).

Sample	**Replicate**	**Nº raw reads (million)**	**Nº collapsed reads (million)**	**Nº reads left (%)**	**Reads ratio *(TAG/CTL)***
Nanos2^{CTL}	1	0.77	0.08	4.2	-
Nanos2^{CTL}	2	82	7.3	3.5	-
Nanos2^{TAG}	1	134	12.6	4.8	158
Nanos2^{TAG}	2	210	28.5	7.7	3.9

Table 3.1. **Number of sequencing reads obtained from CRAC.** The table shows numbers (Nº) of raw and collapsed reads, generated by FastQC Report. Ratio between *Nanos2*^{TAG} and *Nanos2*^{CTL} reads are also reported in the last column of the table.

Dr. Turowski performed the initial reads processing and the bioinformatic analysis. First of all, in-line barcodes, which are specific for each sample, were removed from raw reads. Additionally, in order to remove PCR duplicates,

reads with the same UMI were collapsed to single reads. These reads were also subjected to quality-filtering, before aligning them to the mouse genome. Importantly, *Nanos2*^{TAG} replicates had more reads than *Nanos2*^{CTL}. Moreover, the ratio of *Nanos2*^{TAG} to *Nanos2*^{CTL} reads showed a 158-fold and 3.9-fold enrichment for the first and the second replicate, respectively (last column, Table 3.1). Overall, this indicated a successful enrichment of RNA in *Nanos2*^{TAG} samples compared to controls.

Pre-processed reads were aligned to the mouse genome (mm10), by using different alignment programs. First of all, we assessed the degree of similarity among replicates by measuring the Spearman correlation coefficients (ρ). In Figure 3.8, correlation is displayed as a heatmap, and ρ values are provided on the right, with ρ=1 meaning perfect correlation and ρ=0 no correlation (anti-correlation is not shown). Strikingly, *Nanos2*^{TAG} replicates (1 and 2) were highly correlated between each other (ρ~0.8). Moreover, *Nanos2*^{TAG} samples showed no correlation with *Nanos2*^{CTL}, indicating the high specificity of the CRAC experiments. Eventually, *Nanos2*^{CTL} replicates poorly correlated with each other (ρ~0.3), suggesting that only non-specific RNAs were purified from these samples (negative controls).

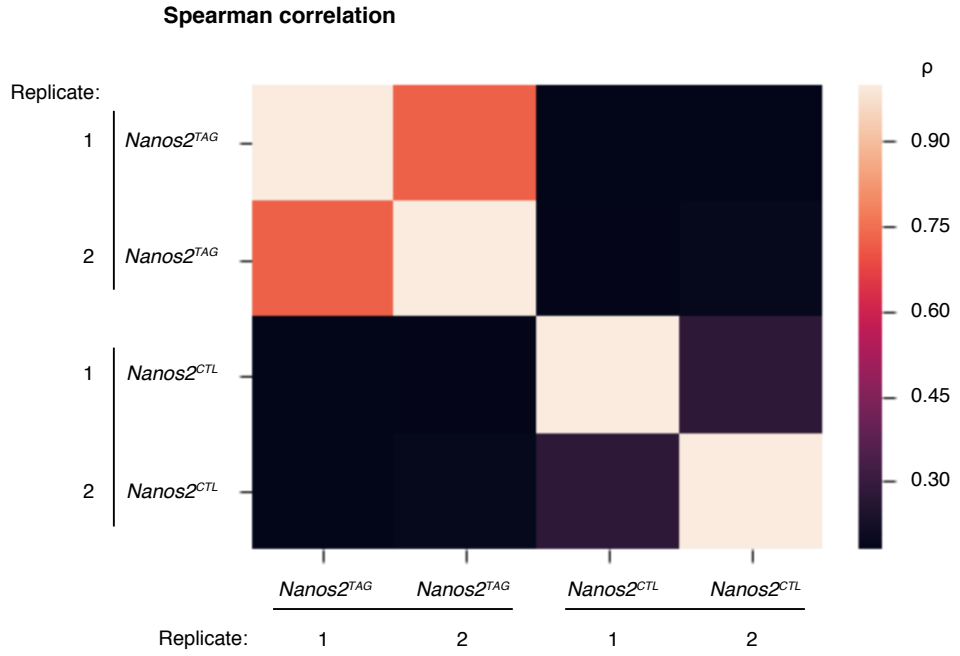


Figure 3.8 Heatmap showing Spearman correlation coefficients (ρ) for CRAC samples.

Heatmap showing the level of correlation among *Nanos2*^{CTL} and *Nanos2*^{TAG} reads and among different replicates (1 and 2). Spearman correlation coefficients (ρ) are shown on the right. ρ=1 means perfect positive correlation and it is in faint pink; ρ=0 means no correlation and is shown in black.

3.7.2 Analysis of the genomic distribution of CRAC reads

After evaluating the similarity among different CRAC samples, we sought to determine the identity of the RNAs bound by NANOS2. To do so, we analysed the genomic distribution of CRAC reads among different classes of RNA (eg. rRNA, mRNA, snRNA, etcetera) (Figure 3.9). Strikingly, both *Nanos2*^{TAG} replicates showed that the vast majority of NANOS2 targets were constituted by protein-coding genes (88.8%).

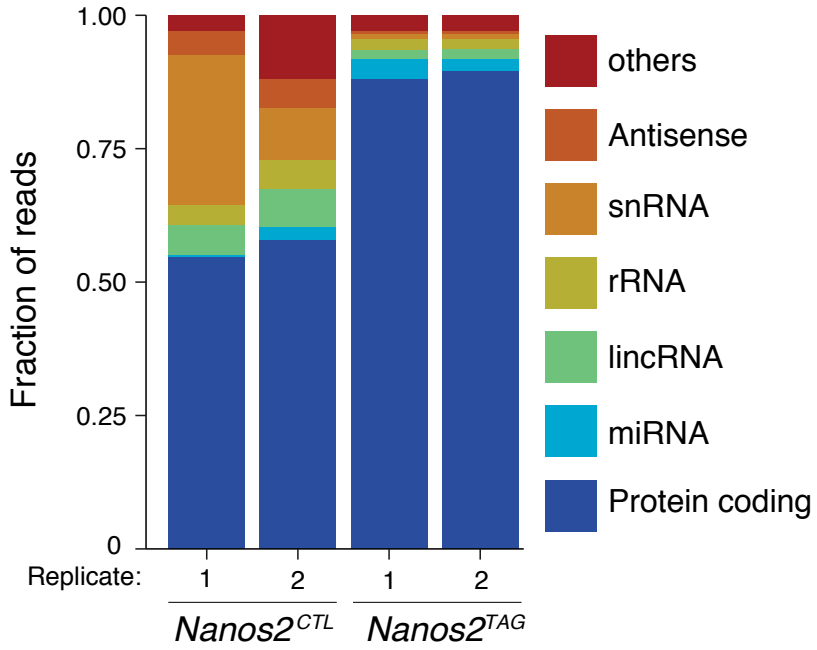


Figure 3.9 Genomic distribution of NANOS2 CRAC reads among different classes of RNA.

Bar graph showing the distribution of CRAC mapped reads among various RNA classes, listed on the right. Reads come from *Nanos2*^{CTL} and *Nanos2*^{TAG} samples. Both replicates (1 and 2) are shown. Reads have not been normalised.

This class was followed by miRNA (3.06%), rRNA (1.885%), lincRNA (1.8%), snRNA (0.965%), antisense (0.475%), and others (2.98%). The class 'others' included mainly processed pseudogenes, but also: transcribed processed pseudogenes, miscRNA, processed transcripts, sense intronic, Mt rRNA, unprocessed pseudogenes, snoRNA, transcribed unprocessed pseudogenes (see the Ensembl database for the exact definition of each category).

Thus, these data revealed for the first time that NANOS2 also binds to non-coding RNAs (ncRNA), and not only to mRNAs, although to a small extent.

Importantly, *Nanos2*^{CTL} and *Nanos2*^{TAG} displayed distinct genomic distribution patterns, and *Nanos2*^{TAG} had considerably depleted abundant classes of RNA,

such as rRNA and snRNA, in both replicates. Overall, this highlighted again the high specificity of the CRAC experiments.

Furthermore, by analysing the list of NANOS2 CRAC targets (ranked according to the number of mapped reads), I noticed that miRNAs were often found among the highest positions. In Table 3.2, I listed the majority of miRNA targets and annotated their CRAC ranking position. Additionally, while examining this list I noticed that many miRNAs were located on the X-chromosome (highlighted in red in Table 3.2). Then, I analysed the NANOS2 CRAC signal associated to these X-linked miRNAs, by using the IGV (Integrative Genomics Viewer) (Robinson et al., 2011), and I noticed that many were located in the same genomic region (Figure 3.10). Interestingly, this miRNA-rich locus has also been described by other authors (Zhang et al., 2019), who performed small RNA sequencing from different spermatogenic populations of adult mouse testis. The expression of these miRNAs was shown to be testis-specific, and the locus was found to be highly conserved across mammals, including human. Therefore, these miRNAs have been called ‘spermatogenesis-related miRNAs’ or ‘spermiRs’. Figure 3.10 shows a snapshot of this X-linked locus with the spermiRs and their associated NANOS2 CRAC signal.

In summary, these data suggest that miRNAs constitute a novel class of NANOS2 targets, which also includes a highly conserved X-linked miRNA family, called spermiRs .

CRAC Ranking	Average number of mapped reads	Gene_symbol
7	22123	Mir871
11	16041.5	Mir741
24	10869	Mir881
31	9132	Mir471
42	8362.5	Mir26a-1
43	8238.5	Mir26a-2
47	7946.5	Mir465c-2
49	7933	Mir465c-1
53	7620	Mir17hg
67	6510	Mir465
80	5962.5	Mir16-2
81	5957	Mir27a
82	5936.5	Mir16-1
83	5904.5	Mir465b-2
87	5855.5	Mir465b-1
89	5801	Mir26b
95	5639	Mir743
172	4464	Mir880
173	4455	Mir6236
193	4261.5	Mir25
260	3714.5	Mir29a
279	3581	Mir31
284	3546.5	Mir883a
315	3379	Mir23a
336	3266.5	Mir7-1
342	3235.5	Mir467d
349	3209.5	Mir743b
383	3073.5	Mir15a
421	2921.5	Mir467e
431	2878	Mir878
439	2851	Mir470
446	2831	Mir1983
461	2786	Mir7-2
478	2737	Mir10a
558	2513.5	Mir23b
573	2478.5	Mir182
581	2460.5	Mir148a
666	2283	Mir146b
705	2213.5	Mir463
737	2142.5	Mir125b-2

CRAC Ranking	Average number of mapped reads	Gene_symbol
798	2039	Mir32
854	1967	Mir6240
884	1936.5	Mir27b
921	1893	Mir20a
923	1889	Mir5099
986	1821.5	Mir125a
1069	1743	Mir19b-1
1097	1712	Mir30e
1114	1698.5	Mir19b-2
1119	1693.5	Mir30f
1134	1681	Mir883b
1149	1671.5	Mir15b
1280	1573	Mir195a
1317	1543.5	Mir92-1
1377	1505	Mir423
1774	1249	Mir30b
1800	1237.5	Mir181c
1855	1208	Mir106b
1897	1184.5	Mir674
1959	1159.5	Mir17
2136	1088.5	Mir199b
2225	1055.5	Mir296
2342	1014.5	Mir350
2393	995	Mir467b
2411	990.5	Mir93
2515	960.5	Mir29b-1
2568	949	Mir101b
2672	917	Mir28a
2692	911.5	Mir467a-8
2718	905	Mir742
2766	894.5	Mir467a-10
2769	894	Mir467a-9
2791	889	Mir467a-3
2857	875.5	Mir297c
2863	873	Mir143
2872	871	Mir467a-1
2881	869.5	Mir467a-7
2897	866.5	Mir21a
2914	862.5	Mir467a-4
2940	856.5	Mir467a-5

Table 3.2 List of miRNAs present among NANOS2 CRAC targets.

miRNAs in the NANOS2 CRAC list are ranked according to their average number of mapped reads, obtained from two CRAC. MiRNA targets which are on the X-chromosome are highlighted in red, whereas miRNAs complementary to Mcm7 on chromosome 5, are shown in green.

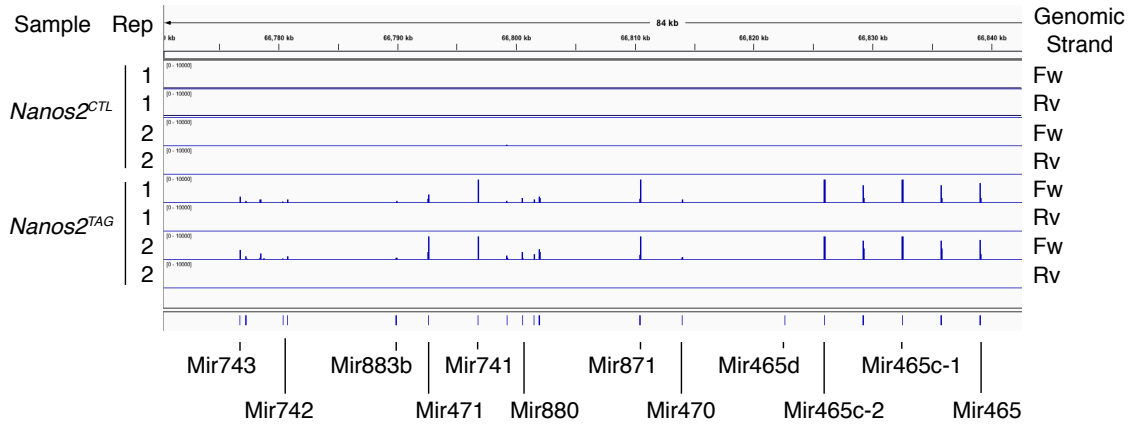


Figure 3.10 Snapshot of an X-linked locus, showing NANOS2 CRAC signal. NANOS2 binds to specific miRNAs (also known as spermiRs), shown in blue. Peaks come from *Nanos2*^{CTL} and *Nanos2*^{TAG} samples, replicates 1 and 2 (Rep). The genomic strand (Fw or Rv) for each sample is also shown on the right. Specific miRNA names are shown on the bottom (eg. Mir743, Mir742 etcetera). NANOS2 CRAC signal (number of mapped reads) was visualised with IGV.

Moreover, by analysing the same list of miRNA targets, I noticed another interesting group of miRNAs. This included miRNA 25, 106b and 93 (highlighted in green in Table 3.2), which come from the same genomic locus, on chromosome 5. Interestingly, they all overlap the same gene, Mcm7, which is the mini-chromosome maintenance complex component 7. This gene is highly conserved across eukaryotes, and it is involved in the initiation of DNA replication, cell cycle regulation, and it is over-expressed in many cancers (Blow and Dutta, 2005; Hua et al., 2014; Tian et al., 2017). Importantly, these three miRNAs (25, 106b and 93) are conserved in human, where they also overlap the Mcm7 gene, on chromosome 7. Overall, these data suggest that NANOS2 binds to another group of conserved miRNAs, whose biological relevance in spermatogenesis is still unknown.

3.7.3 Analysis of NANOS2 peaks and normalisation of CRAC signal to mRNA abundance in SSCs

Since mRNA represented by far the most abundant class of NANOS2 targets, we followed up on this category of transcripts. Since NANOS2 is localised in the cytoplasm, we assumed that it bound mainly to mature mRNAs, without introns. Therefore, we decided to align CRAC reads to the mouse transcriptome, and to use this alignment for the rest of the analysis. CRAC reads contained a certain number of deletions, which were induced by the UV cross-linking and reverse transcriptase, during the CRAC protocol (see Chapter 1 and also afterwards). Since these deletions could be useful to detect NANOS2 binding site on RNA, we used them for the motif enrichment analysis. In order to precisely map reads with deletions to the transcriptome, we used an alignment program that had a certain tolerance for gaps (see Material and methods). The resulting alignment showed that approximately 5000 genes had a CRAC signal, expressed as reads per million (rpm). In order to identify the exact genomic location of NANOS2 binding sites, peaks were called for both *Nanos2*^{TAG} replicates. Since *Nanos2*^{CTL} samples had previously shown no positive correlation with *Nanos2*^{TAG}, we did not call peaks for *Nanos2*^{CTL}, as they would not even be useful to control the background noise.
Peak calling for *Nanos2*^{TAG} samples identified a list of genes ranked according to the 'peak score', defined as the number of reads (rpm) associated with all peaks in that gene. Although 'peak scores' for both replicates were very similar, we decided to use their average rpm values. Eventually, ~2000 genes were identified as containing confident NANOS2 CRAC peaks.

So far, NANOS2 binding intensity on RNA was expressed as 'peak score' (rpm). However, the binding intensity between an RBP and its targets depends on at least three factors. First, the structural and chemical properties at the amino acids-ribonucleotides interface; second, the physiological levels of the

RBP in cells; third, the physiological levels of the interacting RNAs, in the same cellular compartment of the RBP.

CRAC signal or 'peak score' can be considered as a measure directly proportional to the first factor, although it does not completely take into account the third one, that is RNA abundance in cells.

Therefore, to obtain a more precise estimation of NANOS2 binding intensity to its targets, we also decided to quantify mRNA abundance in SSCs. To this end, we performed RNA-seq from two biological replicates, constituted by *Nanos2*$^{TAG/TAG}$ and *Nanos2*$^{TAG/+}$ SSC lines. SSCs were grown until confluency and purified from MEF co-culture, as explained in Material and methods. mRNA-seq libraries were prepared and subjected to high-throughput sequencing. Eleven million reads were mapped to the mouse genome for each replicate, and fpkm (fragments per kilobase million) values for both replicates were calculated. Only average fpkm values from both replicates were considered for further analysis.

Importantly, the correlation between CRAC 'peak scores' (rpm) and signal from RNA-seq (fpkm) was shown to be weak, although positive (0.042). This suggested that CRAC signal was mainly affected by NANOS2 binding intensity, rather than by transcript abundance.

Then, we proceeded with the normalisation of CRAC signal over mRNA levels in SSCs, by dividing 'peak scores' (rpm) by the average fpkm. Values obtained from this ratio were used to rank again CRAC peaks, which generated a new list of normalised NANOS2 CRAC targets. This new list, called "normalised CRAC list" from now onwards, was used for all the remaining analysis.

After this normalisation, the number of NANOS2 peaks was reduced from ~2000 to 1428 genes. This decrease was due to the exclusion of many genes, as they were not expressed at sufficient levels in SSCs (fpkm~0).

3.7.4 Transcriptomic features of the normalised NANOS2 peaks

To explore NANOS2 binding pattern across its targets, we examined how many times NANOS2 bound each transcript. Strikingly, we found that the majority of NANOS2 targets had a single peak, whereas smaller proportions of targets displayed two, three, four, or five peaks per transcript (Figure 3.11).

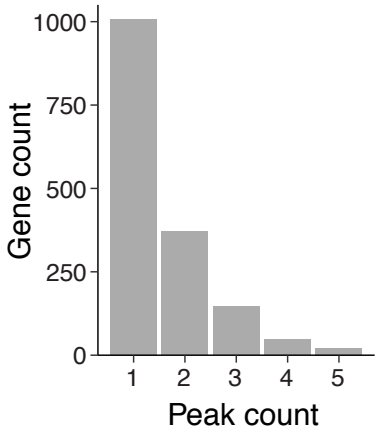


Figure 3.11 Number of NANOS2 CRAC peaks per transcript.
The bar graph shows the distribution of NANOS2 peaks among its targets. In particular, it shows the number of targets which has only 1 or more (2, 3, 4, or 5) peaks per transcript (peak count).

Furthermore, we inspected the shape of NANOS2 peaks in the IGV (Integrative Genomics Viewer). Representative examples of NANOS2 peaks are shown in Figure 3.12. Interestingly, the highest peaks present in these genes are always located in the last exon, which includes the 3' UTR.
Although sometimes multiple peaks per gene are present, these have generally an extremely low signal compared to the main peak.

Therefore, this suggested that the majority of transcripts with multiple peaks contains only one sharp peak and other very minor ones.

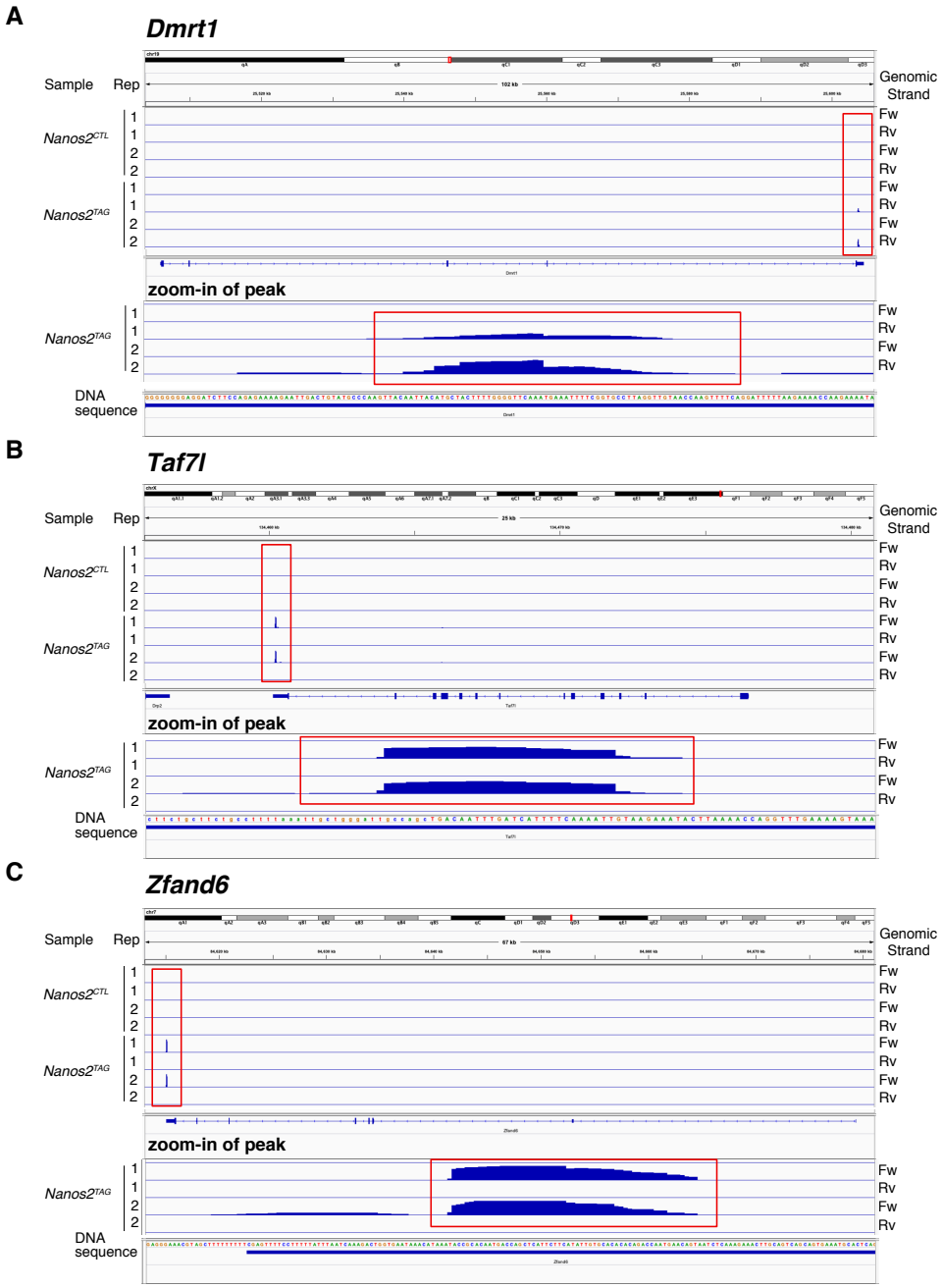


Figure 3.12 Representative examples of NANOS2 CRAC peaks in three targets. The number of mapped reads from *Nanos2*^{CTL} and *Nanos2*^{TAG} CRAC samples (replicates 1 and 2 (Rep)) was used to visualise NANOS2 peaks in IGV. Only three

representative targets are displayed in the picture: *Dmrt1* (A)*, Taf7l* (B)*, and Zfand6* (C). The genomic strand (Fw or Rv) for each sample is shown on the right. Peaks' positions are also shown in relation to exons and introns.

Moreover, to better understand how NANOS2 regulated its targets, we mapped peaks according to different regions of protein-coding transcripts: 5' UTR, CDS (coding sequence) and 3' UTR. These regions are known to be functionally different, and they associate to distinct RBPs and regulatory small RNAs which, in turn, determine the fate of that transcript (see Chapter 1). Remarkably, the vast majority of NANOS2 peaks mapped to 3' UTR, (81.1%), whereas only a small portion mapped to CDS (17.9%), and almost none of them was found in 5' UTR (1.1%) (Figure 3.13). Therefore, NANOS2 binding mainly occurred in the 3' UTR of protein-coding genes, which is a critical region for the regulation of mRNA stability and translation. Overall, these data suggest that NANOS2 might exert a key role in the post-transcriptional regulation of its mRNA targets, in SSCs.

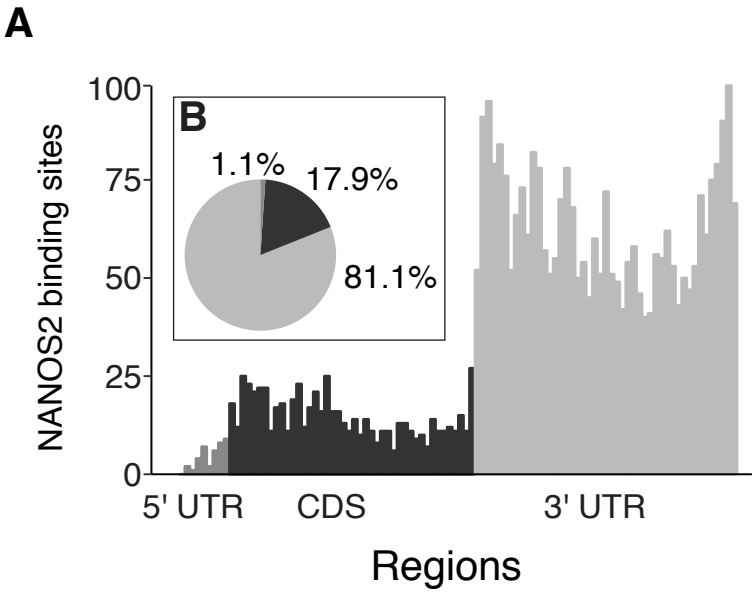


Figure 3.13 Metaprofile of NANOS2 binding sites along the 5' UTR, CDS, and 3' UTR of mRNA targets.

A. The graph shows NANOS2 binding profile along the three regulatory regions of mRNA: 5' UTR, CDS, and 3' UTR. In the x-axis the lengths of each region correspond to the average lengths of all 5' UTRs, CDSs, 3' UTRs used in the analysis. **B.** Relative proportions (percentages) of NANOS2 peaks within the three regions are also reported in the pie chart.

3.7.5 Characterisation of NANOS2 direct binding site using reads with deletions and *de novo* motif analysis

To gain insight into how NANOS2 selects its targets in SSCs, we sought to determine whether it bound RNA with sequence specificity.

To that end, we performed *de novo* motif analysis by using MEME-ChIP (Bailey et al., 2009). As input for the analysis, we used both normalised and non-normalised NANOS2 peaks (see the previous part) and compared the results. Moreover, we also took advantage of sequencing reads containing deletions. Indeed, deletions can be used to precisely map the binding site of an RBP on RNA, at single-nucleotide resolution (see Chapter 1).

Therefore, NANOS2 CRAC reads were filtered in order to select the ones with deletions. These reads were centred on deletions, and only one hundred or twenty nucleotides in the proximity of deletions were further analysed.
We used reads with and without deletions to perform motif enrichment analysis. We also used either all NANOS2 peaks or the highest peak per gene. Additionally, we selected different subsets of genes from both the normalised and non-normalised CRAC lists and compared the results. Interestingly, a common motif appeared to be enriched in CRAC targets, in almost all the cases described above.
Here, I show only one example of the various analyses performed.
In this case, we selected the first half of the normalised CRAC peaks (714 out of 1428 peaks), which corresponded to 5530 deletions. Also, we selected only twenty nucleotides around deletions. As a negative control, we used the same sequences, but shuffled. Interestingly, MEME-ChIP showed that an AU-rich heptamer, 'AUNAANU', was enriched with high confidence (P-value =1.5 e^{-067}) in 970 deletions, one-fifth of the total (Figure 3.14 A). Since we considered only twenty nucleotides around deletions, we estimated that actually more reads were likely to contain this motif, downstream from the cross-linked nucleotides. Visual inspection of these reads revealed that one or more uridine residues (U) were often present in the centre, at positions corresponding to the cross-linked nucleotides or deletions. Remarkably, a uridine residue stood out in the centre of the motif
Overall, this suggested that uridine residues were the most probable ribonucleotides directly interacting with NANOS2.

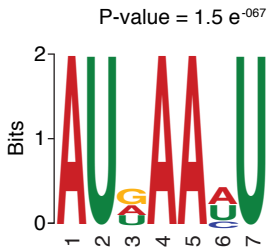
Examples of NANOS2 targets containing the motif adjacent to deletions (uridines) are shown in Figure 3.14 B.
As mentioned before, the 'AUNAANU' motif was repeatedly found in many of the analyses performed. However, when analysing the top two or three hundred targets, from the normalised CRAC list, the motif became even more

U-rich, with the third nucleotide changing from N to U, 'AUUAANU' (data not shown).

Strikingly, this motif was not found in the ATtRACT database (Giudice et al., 2016), which contains all known RBP motifs. Therefore, the RNA heptamer 'AUNAANU' represented a novel motif, exclusively bound by NANOS2.
Interestingly, MEME-ChIP did not detect additional motifs with comparable P values, following this one. This suggested that NANOS2 targets did not contain other known RBPs motifs, in the proximity of NANOS2 binding site.
Importantly, the enrichment of the 'AUNAANU' motif among CRAC targets did not correlate with CRAC signal intensity. This suggested that additional factors, other than sequence specificity, could also influence targets' selection and determine the intensity of NANOS2 binding to RNA.

In summary, these data indicated that NANOS2 binds RNAs enriched for the novel 'AUNAANU' motif, and therefore this RBP displays sequence specificity in mouse SSCs.

A



B

Genes	Motif AUNAANU near deletion (U)
Ythdf2	AGGAUUAAUUUGAUUUCAAA
Taf7l	AAAUCAACUUUAGUGAGUAC
Zfand6	CACCAGUCUUUGAUUAAAUA
Otud4	GAUGAAUUUUUGGCAUUGUG
Rpl13	CAAUAAAUUUUCCAUAAAGC
Slc30a1	UAAUGAAAUUUGAUUCCUGG
Zbtb10	AUGAAAUCUUUAUAGUCUCC
Klf11	AUAUUAAUUUUACUUUUCUG
Yy1	UUAAUCAAUUUAAACCCAUU
Sycp1	UAAGAAUACUGAAUUAACUG

Figure 3.14. *De novo* motif analysis using NANOS2 CRAC targets.

A. CRAC reads containing deletions (cross-linking sites) were analysed by performing *de novo* motif analysis with MEME-ChIP. The most enriched motif, 'AUNAANU', is

shown as a logo in the figure, with the corresponding P-value. **B.** Sequence alignment of a subset of NANOS2 reads containing the 'AUNAANU' motif (in red). Reads were centered on deletions (central 'U', in bold red), which correspond to the most probable cross-linking sites (and direct NANOS2 binding site).

3.7.6 *In silico* prediction of the secondary structure of the NANOS2 motif sequence, using *Taf7l* 3' UTR

So far, I have shown that NANOS2 specifically binds the 3' UTRs of mRNAs, enriched in the novel 'AUNAANU' motif. Additionally, this motif was often located near some uridine bases, which represented most likely the direct sites of interaction between NANOS2 and RNA.

It is reasonable to speculate, that the 'AUNAANU' sequence need to be located in an accessible region of the RNA, in order to allow NANOS2 binding. To address this hypothesis, I sought to understand the spatial conformation of NANOS2 binding sites, by taking into account also regions in the proximity of the motif. Thus, I choose one of the NANOS2 targets containing the motif, *Taf7l*, and selected its 3' UTR sequence. Then, I performed a preliminary analysis of its secondary structure by using the 'RNAfold web server' (Gruber et al., 2008; Hofacker, 2003). As shown in Figure 3.15, two possible secondary structures are predicted for this 3' UTR. In 'A.' it is shown the minimum free energy structure (MFE), and in 'B.' the centroid structure (Ding et al., 2005), which is an alternative way to model RNA structures. Apart from the central region, both algorithms show a similar conformation for *Taf7l* 3' UTR, which makes the overall prediction more reliable. Then, to assess the local conformation in the close proximity of NANOS2 motif, I first identified the 'AUNAANU' sequence within both structures, (arrows in Figure 3.15). It is worth noting that in the drawing, each base has a colour that reflects the probability of base-pairing, with blue representing the minimum base-pair probability (0), and red the maximum one (1). Strikingly, in both structures,

most of the nucleotides included in the NANOS2 motif are not engaged in strong base pairing (they are coloured in green or light blue). Indeed, these bases reside within one or two loops, depending on the MFE or centroid structure. Overall, this suggested that the RNA bases of the NANOS2 motif are in a relatively open conformation, which is not highly structured. Importantly, this might increase the chances of NANOS2 binding to the 'AUNAANU' sequence.

However, further analysis will be necessary to confirm that this type of secondary structure is also present in the rest of NANOS2 motifs, from the other targets.

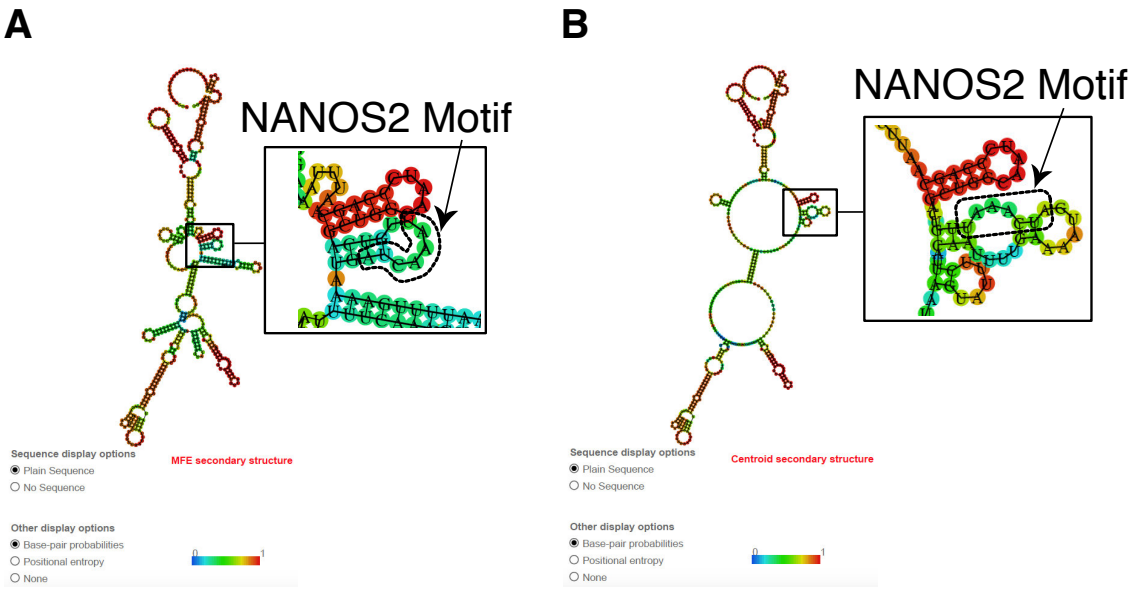


Figure 3.15. Prediction of *Taf7l* 3' UTR secondary structure, using the 'RNAfold web server'.

A. MFE (minimum free energy) secondary structure of *Taf7l* 3' UTR. On the right, zoom-in of the region containing the NANOS2 motif (arrow). **B.** Centroid secondary structure of *Taf7l* 3' UTR and zoom-in of the nucleotides corresponding to the NANOS2 motif.

3.7.7 Identity of NANOS2 targets and Gene ontology analysis

In order to better understand how NANOS2 maintains SSCs, we sought to relate NANOS2 CRAC targets to their biological function. Thus, I searched for the biological processes enriched in NANOS2 targets, by using the g:Profiler software (Reimand et al., 2019). The normalised list of CRAC targets, including 1428 genes, was provided to the software, and results were subsequently analysed. As shown in Figure 3.16 A, NANOS2 targets were mainly enriched in cellular metabolic processes which concerned “cellular macromolecules”, “nitrogen compounds”, “proteins”, “organic substances”, “nucleic acids”, “nucleobase-containing compounds”, “RNA”, “heterocycle”, “cellular aromatic compounds” as well as “regulation of metabolism”. Besides these metabolic processes, also “organelle organisation” and “gene expression” were found to be enriched.

I noticed that the majority of these processes belonged to broad biological categories, and additionally, they were closely related to each other, suggesting a certain degree of redundancy.

In order to simplify the interpretation of GO results, and to potentially discover more specific biological functions, I reduced the GO term size from 10000 to 1000. Results are shown in Figure 3.16 B. Interestingly, this analysis revealed enrichment for additional processes, such as “protein modification by small protein conjugation”, “translation”, “chromatin organisation”, “protein ubiquitination”, and “regulation of translation”. However, they also confirmed a substantial enrichment for many biosynthetic processes.

A

Term size = 10,000

GO:BP		stats	
Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$
cellular macromolecule metabolic process	GO:0044260	8.714×10^{-46}	
cellular metabolic process	GO:0044237	1.934×10^{-45}	
nitrogen compound metabolic process	GO:0006807	3.104×10^{-40}	
cellular nitrogen compound metabolic process	GO:0034641	1.088×10^{-39}	
macromolecule metabolic process	GO:0043170	6.319×10^{-39}	
primary metabolic process	GO:0044238	1.997×10^{-38}	
cellular protein metabolic process	GO:0044267	6.388×10^{-36}	
organelle organization	GO:0006996	3.230×10^{-35}	
macromolecule biosynthetic process	GO:0009059	4.445×10^{-35}	
gene expression	GO:0010467	6.271×10^{-35}	
cellular macromolecule biosynthetic process	GO:0034645	1.129×10^{-34}	
organic substance metabolic process	GO:0071704	1.138×10^{-33}	
nucleic acid metabolic process	GO:0090304	1.321×10^{-32}	
cellular nitrogen compound biosynthetic process	GO:0044271	2.114×10^{-32}	
regulation of cellular metabolic process	GO:0031323	2.333×10^{-32}	
nucleobase-containing compound metabolic process	GO:0006139	3.815×10^{-32}	
biosynthetic process	GO:0009058	5.166×10^{-32}	
organic substance biosynthetic process	GO:1901576	1.578×10^{-31}	
metabolic process	GO:0008152	2.678×10^{-31}	
regulation of metabolic process	GO:0019222	2.885×10^{-31}	
cellular biosynthetic process	GO:0044249	3.853×10^{-31}	
RNA metabolic process	GO:0016070	7.547×10^{-30}	
heterocycle metabolic process	GO:0046483	9.493×10^{-30}	
cellular aromatic compound metabolic process	GO:0006725	1.061×10^{-29}	
regulation of macromolecule metabolic process	GO:0060255	3.175×10^{-29}	

B

Term size = 1,000

GO:BP		stats	
Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$
protein modification by small protein conjugation or rem...	GO:0070647	5.034×10^{-23}	
peptide biosynthetic process	GO:0043043	7.258×10^{-23}	
translation	GO:0006412	2.082×10^{-22}	
protein modification by small protein conjugation	GO:0032446	4.717×10^{-21}	
chromatin organization	GO:0006325	6.463×10^{-20}	
amide biosynthetic process	GO:0043604	1.594×10^{-19}	
cellular macromolecule catabolic process	GO:0044265	1.684×10^{-19}	
peptide metabolic process	GO:0006518	6.274×10^{-19}	
protein ubiquitination	GO:0016567	9.175×10^{-19}	
mRNA metabolic process	GO:0016071	1.329×10^{-16}	
regulation of translation	GO:0006417	2.146×10^{-16}	
covalent chromatin modification	GO:0016569	1.611×10^{-15}	
regulation of cellular amide metabolic process	GO:0034248	1.634×10^{-15}	
peptidyl-lysine modification	GO:0018205	1.794×10^{-15}	
histone modification	GO:0016570	1.996×10^{-14}	
intracellular protein transport	GO:0006886	2.762×10^{-14}	
cellular protein-containing complex assembly	GO:0034622	4.080×10^{-14}	
cellular amide metabolic process	GO:0043603	4.185×10^{-14}	
negative regulation of translation	GO:0017148	1.202×10^{-12}	
ribonucleoprotein complex biogenesis	GO:0022613	1.211×10^{-12}	
protein localization to organelle	GO:0033365	1.557×10^{-12}	
regulation of mRNA metabolic process	GO:1903311	1.792×10^{-12}	
protein catabolic process	GO:0030163	9.337×10^{-12}	
posttranscriptional regulation of gene expression	GO:0010608	1.045×10^{-11}	
negative regulation of cellular amide metabolic process	GO:0034249	1.432×10^{-11}	
negative regulation of transcription by RNA polymerase II	GO:0000122	1.585×10^{-11}	
ribonucleoprotein complex assembly	GO:0022618	2.191×10^{-11}	

Figure 3.16. Gene ontology (GO) analysis performed on ranked NANOS2 CRAC targets. GO performed with term size of 10,000 (A.), or 1,000, (B.).

Therefore, NANOS2 mainly binds to transcripts involved in macromolecules metabolism, including nucleic acids and proteins, and more specifically nitrogen and aromatic compounds. Additionally, NANOS2 binds to regulators of cellular metabolism, probably to ensure a more robust control on this cellular process.
However, NANOS2 targets are also involved in gene expression at multiple levels, including chromatin organisation, translation, and its regulation, as well as post-translational modifications, such as ubiquitination.
Overall, these results indicate that NANOS2 might have an important role in controlling SSC growth, through the regulation of major biosynthetic processes. Furthermore, by binding to mRNAs which regulate gene expression at multiple layers, NANOS2 could shape SSC identity and thus pave the way for either self-renewal or differentiation.

Tight regulation of cellular metabolism is known to be essential for SSCs. In particular, the repression of the mTOR pathway is critical to maintain the self-renewing SSC pool over time (Hobbs et al., 2010; Wang et al., 2016). NANOS2 was related to mTORC1 by virtue of their physical protein-protein interaction, which leads to mTORC1 trapping in stress granules and reduces translation levels in SSC lines. This mechanism is also thought to contribute to the repression of NANOS2 targets (Zhou et al., 2015). To further explore the relation between NANOS2 and the mTOR pathway, I assessed whether NANOS2 interacted with mTorc1 mRNA and mTOR regulators (defined as in MGI informatics, GO terms database). Although I did not detect mTorc1 among CRAC targets, strikingly, I could identify eleven positive mTOR regulators out of thirty-nine, in our CRAC list. Interestingly, only three negative mTOR regulators out of forty-six were recruited by NANOS2, and they were bound only with a moderate intensity (none of them was found within the first 500 CRAC targets). In contrast, five positive mTOR regulators out of eleven were present among the top 500 CRAC targets, and they were: Lamtor3, Rheb, Xbp1, Rictor and Mios. Overall, these findings suggest that NANOS2 might

regulate mTOR signalling through direct binding of its positive regulators. This could potentially constitute an additional mechanism through which NANOS2 keeps under control mTOR pathway in SSCs, in order to preserve their self-renewal.

Afterwards, I assessed whether known NANOS2 targets, previously determined by RIP in SSCs (Zhou et al., 2015) and embryonic gonads (Saba et al., 2014), were also identified by CRAC. Among the most confident RIP candidates, there were *Taf7l*, *Sohlh2*, *Dmrt1*, and *Dazl*, all involved in germ cell differentiation (see Chapter 1). Interestingly, I found that only *Taf7l* and *Dmrt1* were confirmed by CRAC, indicating that *Sohlh2* and *Dazl* are not directly bound by NANOS2 in SSC lines.

In summary, I have shown that NANOS2 targets are enriched for cellular metabolic processes and that they also include positive mTOR regulators.
Therefore, these data suggest that upon direct recruitment of various metabolism-related transcripts, NANOS2 might control SSC growth and thus preserve the self-renewing SSC pool for a long time.

3.8 Analysis of the transcriptional changes through spermatogonia development and comparison with NANOS2 CRAC targets

In order to better understand the biological function of NANOS2 targets in SSCs, we sought to analyse their expression across SSC development, in the adult testis. Therefore, we characterised the transcriptomic profile of three different stages of spermatogonia by performing RNA-seq. The first stage we analysed was the self-renewing SSC pool, marked by Gfrα1. The second population of spermatogonia was represented by transit-amplifying cells, which express Miwi2 (Carrieri et al., 2017). The third stage was constituted by differentiated spermatogonia, marked by both Miwi2 and c-kit. Importantly, most of the Nanos2 positive cells also express Gfrα1, as determined by

previous studies and also our IF data. Therefore, the first stage of spermatogonia that we analysed, was also the one closest to Nanos2 expressing cells.

In order to isolate these three populations from adult mouse testes, we employed two different transcriptional reporter alleles, available in our laboratory: a *Gfrα1*^{GFP} allele (Hara et al., 2014), where the gene encoding the GFP protein was inserted into the Gfrα1 locus, and a *Miwi2*^{TOM} allele (Carrieri et al., 2017), where the gene encoding the Tomato protein was inserted into the first exon of *Miwi2*. These two alleles have been crossed to generate a *Gfrα1*$^{GFP/+}$; *Miwi2*$^{TOM/+}$ line, which was used for FACS. Spermatogonia have been isolated from testes of *Gfrα1*$^{GFP/+}$; *Miwi2*$^{TOM/+}$ mice by FACS, whose gating strategy is shown in Figure 3.17. The first two stages of spermatogonia were selected by gating the single and live cells. Somatic cells were gated out by using the CD45 and CD51 markers (CD45^{neg} CD51^{neg}, live cells). Then, to select only undifferentiated spermatogonia, we sorted cells positive for the surface marker CD9, and we gated out c-kit-positive cells (c-kit^{neg}; CD9^{pos} cells). Next, we analysed the GFP fluorescent signal, which indicated the expression of Gfrα1^{+} SSCs, and the Tomato fluorescence, for the Miwi2-Tom^{+} cells. Thus, we isolated the GFP^{pos}, TOM^{neg} cells, which represented the Gfrα1^{+} cells and, at the same time, we sorted the GFP^{neg}, TOM^{pos} cells, which are the Miwi2-Tom^{+} cells (Figure 3.17 A). Additionally, by using other mice, we also sorted the third stage of spermatogonia, by gating the live, CD45^{neg} CD51^{neg} cells, and selecting the Miwi2-Tom^{+} c-kit^{+} cells (Figure 3.17 B).

RNA was isolated from these populations in order to generate mRNA libraries for sequencing. FACS and RNA isolation have been performed by Dr. Ivanova, whereas the subsequent bioinformatic analysis was performed by Dr. Enright and his group.

First of all, we analysed the similarity among the four biological replicates (Figure 3.18). Importantly, principal component analysis (PCA) showed that our samples clustered only according to the spermatogonial stage, which was either Gfrα1^{+} (green circles), or Miwi2-Tom^{+} (red triangles), or Miwi-Tom^{+} c-kit^{+}

(blue squares). Therefore, the four replicates displayed very low biological variability.

Then, we characterised gene expression patterns across the three populations of spermatogonia. To this end, we used the Markov clustering algorithm (Freeman et al., 2007), which identifies clusters of genes whose expression profile varies across different developmental stages .

This analysis revealed that twenty-two clusters of genes had different profiles across spermatogonia development. The full list of clusters, with the corresponding number of genes per cluster, is shown in Table 3.3, whereas the expression dynamics of the first ten clusters are shown in Figure 3.19.

Cluster 1 was the most remarkable group of genes undergoing transcriptional changes across the three developmental stages and included more than six thousand genes. Interestingly, this cluster represented genes which were expressed at a moderate level in Gfrα1^{+} SSCs, and which were acutely up-regulated during the transition from self-renewing SSCs (Gfrα1^{+}), to transit-amplifying cells (Miwi2-Tom^{+}). Moreover, upon differentiation (Miwi2-Tom^{+} c-kit^{+} cells), the same group of genes underwent a significant down-regulation. Similarly, clusters 5, 7, and 10, displayed analogue dynamics across development. Overall, these four clusters represented the great majority of transcripts identified by the Markov algorithm.

Cluster 2 contained the second largest group of genes, over four thousand. This cluster included genes which were sharply down-regulated from the first to the second stage of spermatogonia, and which continued to decrease their expression level until differentiation (third stage).

Next, we sought to determine whether NANOS2 targets were significantly enriched or depleted in any of the clusters identified. This analysis was performed by Dr. van der Lagemaat.

NANOS2 targets were selected from the normalised CRAC list, which included 1428 genes. Surprisingly, NANOS2 targets were slightly enriched or over-represented only in cluster 1 (P-value < 0.05). Moreover, they were depleted,

or under-represented, only in cluster 2 (P-value < 0.01) (see P-values in red in Figure 3.19). Given the expression profiles of these two clusters, these data indicated that most likely NANOS2 bound to many transcripts which were up-regulated during the transition from Gfrα1$^+$ to Miwi2-Tom$^+$ cells.

In summary, this analysis suggested that most of the transcriptional changes occurring through the three developmental stages analysed, were represented by cluster 1. This group included transcripts whose expression increases from self-renewing SSCs to transit-amplifying cells.
Importantly, NANOS2 targets showed a modest overrepresentation among genes in cluster 1.

Nanos2 expression decreases along the three stages analysed. Thus, these data suggested that NANOS2, upon binding to a group of genes in cluster 1, might repress them in Gfrα1$^+$ cells. Then, once SSCs start to lose NANOS2 (Miwi2-Tom$^+$ cells), these genes would be released from its repression and become up-regulated.
Since these genes are involved in the metabolism of nucleic acids and proteins, these data indicate that NANOS2 might control SSCs growth in Gfrα1$^+$ cells, by preventing excessive proliferation and thus preserving the self-renewing pool over time.

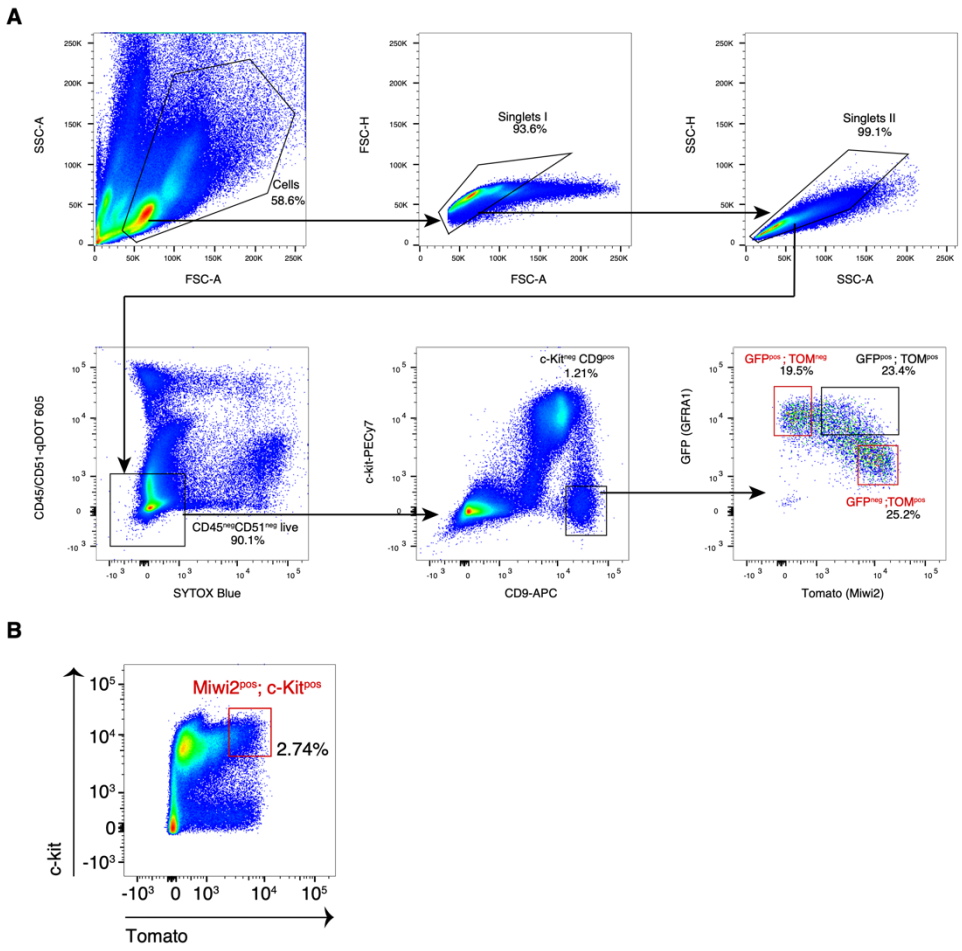


Figure 3.17. Gating strategy used to sort ex-vivo three distinct populations of spermatogonia from adult testes, by FACS.

A. Spermatogonia have been sorted according to size and shape (SSC-A, FCS-A) and cell debris were excluded. Single cells (singlets) were selected according to FSC-H and FSC-A (and also to SSC-H and SSC-A). Somatic cells were gated out by selecting CD45^{neg}; CD51^{neg} cells. Dead cells were also excluded by using a living dye (Sytox Blue). Undifferentiated spermatogonia were selected by gating for c-kit^{neg}; CD9^{pos} cells. SSCs were selected by gating the GFP^{pos}, TOM^{neg} cells (Gfrα1^{+}). Transit-amplifying spermatogonia were selected by gating the GFP^{neg}, TOM^{pos} cells (Miwi2-Tom^{+}). **B.** Representative FACS analysis of Miwi2^{pos}, c-Kit^{pos} cells (Miwi2-Tom^{+} c-kit^{+}), which represent differentiating spermatogonia. These cells are live, CD45^{neg}, CD51^{neg}. The populations used for mRNA-seq and Markov clustering analysis are highlighted in red.

Cluster	Number of genes
1	6064
2	4205
3	2295
4	2037
5	1184
6	1058
7	647
8	640
9	570
10	544
11	282
12	114
13	90
14	85
15	83
16	57
17	19
18	11
19	10
20	1
21	1
22	1

Table 3.3. List of clusters identified across three stages of adult spermatogonia.

The number of genes per cluster is also reported.

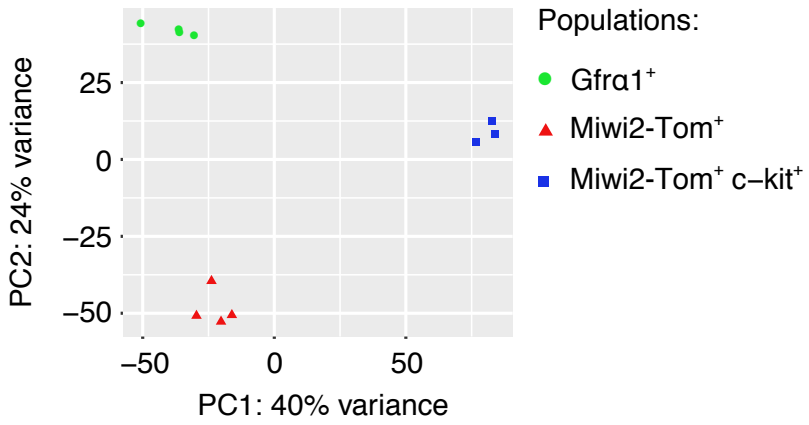


Figure 3.18. Principal component analysis (PCA) of RNA-seq samples from adult spermatogonia.

RNA-seq was performed on three populations of spermatogonia, Gfrα1^{+}, Miwi2-Tom^{+}, and Miwi2-Tom^{+} c-kit^{+}, and four different mice have been used to sort each population. PCA shows the variability among the four biological replicates, and in relation to the spermatogonia stage. PC1, which explained most of the variance of the samples (40%), is shown on the x-axis. PC2 is shown on the y-axis.

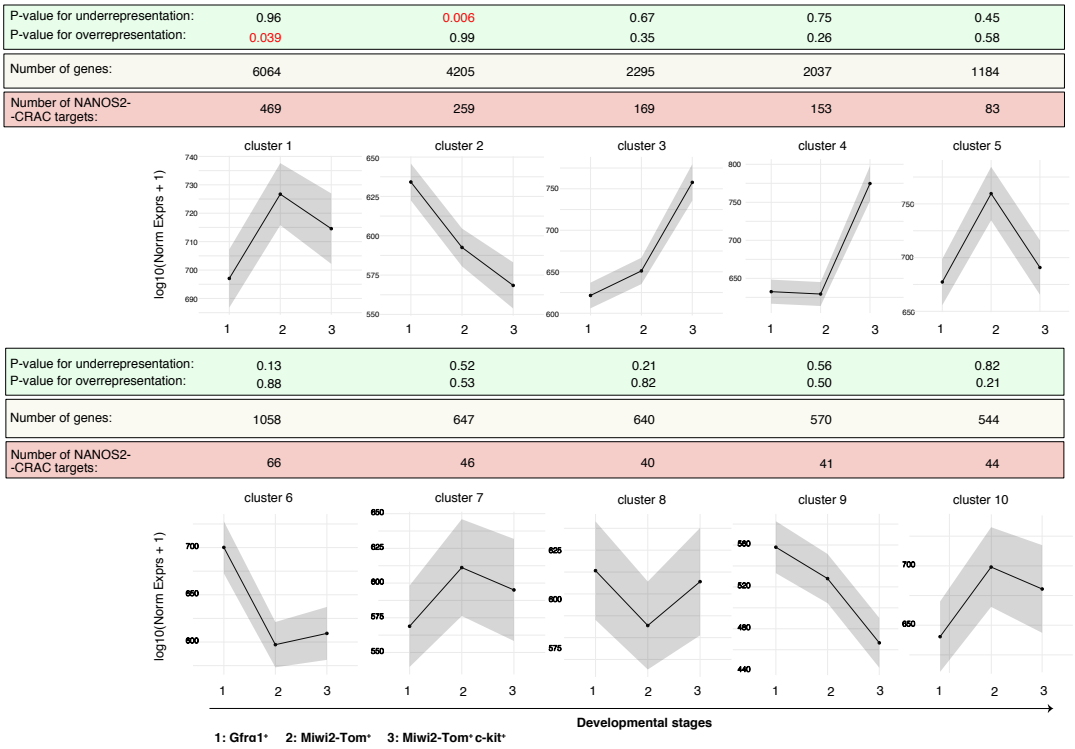


Figure 3.19. Results of Markov clustering analysis performed on three developmental stages of spermatogonia.

The graphs show the expression profile of ten gene clusters, across three populations of spermatogonia: Gfrα1^{pos} (1), Miwi2-Tom^{pos} (2), and Miwi2-Tom^{pos} c-kit^{pos} (3). The number of genes present in each cluster is reported. Additionally, we have investigated whether NANOS2 CRAC targets were significantly over- or under-represented among any of these clusters. The P-values, for both enrichment and depletion, were calculated from the hypergeometric test. Statistically significant P-values are displayed in red ($P < 0.05$).

3.9 Discussion

3.9.1 NANOS2 binds directly to RNA both in HEK293t cells and in SSC lines

In this chapter, I have shown how we developed a *Nanos2*^{TAG} mouse allele, by gene targeting in mouse. *Nanos2*^{TAG} expresses a TAG-NANOS2 fusion protein, whose complex tag allowed performing both biochemical experiments (CRAC) and *in vivo* localisation studies. Importantly, I have shown that the tag did not impair NANOS2 function in male germ cells. Histological analyses of testicular sections from *Nanos2*^{TAG} mice revealed that TAG-NANOS2 could sustain spermatogenesis fully, even in the long term. Moreover, NANOS2 localisation *in vivo* is not disrupted by the tag, as assessed by immunofluorescence of whole-mount seminiferous tubules. Indeed, NANOS2 is correctly expressed in the self-renewing SSCs pool, marked by Gfrα1, and it does not co-localise with c-kit, a marker of differentiated spermatogonia.

To perform CRAC in SSCs, I derived SSC lines from *Nanos2*^{TAG} pups, assessed NANOS2 expression both by WB and by FACS, and expanded SSCs *in vitro*. However, NANOS2 is low abundant in SSC lines, and additionally, these cells divide very slowly. Since CRAC requires high amounts of material, applying it to SSC lines constituted a major challenge. Thus, I first optimised it by over-expressing TAG-NANOS2 in HEK293t cells (*NANOS2*^{OE}), and only later, I performed CRAC in SSC lines.

By using *NANOS2*^{OE} HEK293t cells, I optimised the main steps of the CRAC protocol, which revealed that TAG-NANOS2 was suitable for CRAC. Moreover, by performing the entire CRAC protocol, we discovered that NANOS2 binds directly to RNA in human cell lines. Hence, I used exactly the same conditions to perform CRAC in SSC lines and, importantly, I could confirm that NANOS2 binds directly to RNA also there.

Remarkably, the fact that NANOS2 can bind to RNA in both cell types, strongly suggests that it does not require additional SSCs-specific factors (e.g., RBPs) to bind to RNA. In contrast, NANOS ortholog in *Drosophila* (dNANOS) needs PUMILIO in order to shift RNA, as shown by EMSA (Sonoda and Wharton, 1999; Weidmann et al., 2016). Thus, although the RNA-binding ability of dNANOS was tested only by EMSA and not by CRAC, it seems that the mode of RNA-binding of NANOS orthologs is not fully conserved among different organisms.

Moreover, the strong radioactive signal on the CRAC films indicated that NANOS2 cross-links well to RNA, using the UV light. Indeed, by examining the amino acid sequence of NANOS2 zinc fingers and its proximity, I observed many amino acids which are known to have high cross-linking potential (Shchepachev et al., 2019). These include three aromatic residues, F, Y, H as well as three positively charged, H, R, K and two aliphatic, P and G. Since these amino acids are present in the most conserved region of NANOS2, this suggests that also other NANOS proteins could cross-link well and may be suitable for CRAC. Thus, studying the RNA occupancy of NANOS paralogs and orthologs by CRAC (or CLIP) is potentially possible, in mouse and in different organisms.

3.9.2 NANOS2 binds to the 3' UTR of mRNAs with sequence specificity, in mouse SSC lines

Strikingly, CRAC analysis showed that ~90% of NANOS2 binding occurs on mRNA, in SSC lines. Moreover, after normalising CRAC signal with mRNA abundance, we discovered that NANOS2 binds to ~1400 transcripts, which are almost exclusively constituted by novel targets.
In mouse SSC lines, NANOS2 was already shown to bind to a handful of mRNAs, by native RIP (Suzuki et al., 2010; Zhou et al., 2015). Therefore, our data corroborate the fact that NANOS2 associates with mRNAs and reveal for

the first time that this interaction is direct, and that occurs at the transcriptomic level, in mouse SSC lines.
Moreover, also other NANOS orthologs associated with mRNAs, and these interactions are essential for NANOS functions (see Chapter 1). Thus, not only mRNAs constitute a conserved class of targets for different NANOS, but they could also be the means through which NANOS2 exerts its critical functions, in mouse SSCs.

Furthermore, CRAC analysis showed that the vast majority of NANOS2 binding (~81%) occurs in the 3' UTR of mRNAs, in SSC lines.
NANOS2 was already shown to associate with the 3' UTR of *Dazl* in male gonocytes (Kato et al., 2016), and to the 3' UTR of *Sohlh2*, in SSC lines (Zhou et al., 2015). However, our work determined for the first time that NANOS2 has a substantial bias towards the 3' UTR of its *bona fide* targets.
Given that post-transcriptional gene regulation extensively exploits the 3' UTR of mRNAs (Mayya and Duchaine, 2019), our findings suggest that NANOS2 might directly regulate the expression of many protein-coding genes, at the post-transcriptional level, and through their 3' UTR.
Moreover, previous studies in *Drosophila* have reported that dNANOS binds to the 3' UTR of different mRNAs (Kadyrova et al., 2007; Sonoda and Wharton, 1999; Wharton and Struhl, 1991). Hence, the 3' UTR bias might be a conserved and critical feature of NANOS targets, also among different organisms.

Next, we mapped NANOS2-RNA interactions at the single-nucleotide resolution, by using CRAC reads with deletions, which reflect the cross-linking sites. Strikingly this analysis revealed that NANOS2 targets are enriched in the novel 'AUNAANU' motif. Additionally, this motif is often found adjacent to or in the proximity of the direct binding site, which is constituted by uridine residues (U). Thus, NANOS2 binds to RNA with sequence specificity in mouse SSC lines. Interestingly, the 'AUNAANU' motif is seven nucleotides long, which is

the typical length of the RNA-motifs identified so far, in both human and *Drosophila* (Giudice et al., 2016). It is also likely that this nucleotide length corresponds to the space necessary to accommodate two zinc finger motifs, present in NANOS2 C-terminus. Indeed, it was shown that one zinc finger generally binds to three nucleotides (Choo and Klug, 1994). The fact that we found a heptameric motif instead of a perfect hexamer, it might be explained by the unique structure adopted by NANOS2 zinc fingers. Since these are separated by a particular spacing, not shared with any other known zinc fingers (see Chapter 1), this could justify why NANOS2 binds to seven ribonucleotides instead of six.

So far, nobody had investigated whether NANOS2 bound to RNA with sequence specificity in mouse SSC lines. In *Drosophila*, NANOS is generally thought to bind to RNA without sequence-specificity (Curtis et al., 1997; Hashimoto et al., 2010b). Moreover, its RNA-binding ability depends on PUMILIO, which recruits dNANOS to RNA (Sonoda and Wharton, 1999). However, once dNANOS joins the PUMILIO-RNA complex, it alters PUMILIO sequence specificity, which becomes more AU-rich (Weidmann et al., 2016). Since also the novel NANOS2 motif ('AUNAANU') is AU-rich, this might implicate a certain degree of conservation in the RNA-protein interactions between NANOS proteins and their targets.

As previously mentioned, CRAC analysis revealed that NANOS2 directly binds to uridine residues (U). Interestingly, the crystal structure of PUMILIO-dNANOS-RNA showed that dNANOS mediates base-specific contacts with the last uridine in the 'AUU' sequence (Nanos binding site) (Weidmann et al., 2016). Moreover, the three hydrophobic amino acids, present in dNANOS zinc fingers, which interact with that uridine, are essential to repress *hunchback* mRNA (Weidmann et al., 2016). Therefore, dNANOS-uridine interaction is essential for NANOS function in *Drosophila*. Given the similarity among NANOS proteins, and in particular among their zinc fingers, it is possible that

also the interaction between NANOS2 and the uridine residues detected by CRAC is critical for NANOS2 function, in mouse SSC lines.
However, we cannot exclude that the identification of uridine residues could also be linked to the intrinsic bias of pyrimidines to cross-link (Hockensmith et al., 1986; Meisenheimer and Koch, 1997). Also, analysis of CLIP and iCLIP experiments of different RBPs revealed that the 'UUU' motif is often enriched at the cross-linking sites (Sugimoto et al., 2012). Additionally, some types of reverse transcriptases have been shown to have a very high rate of single-base deletion error, in homopolymeric nucleotides (Hamburgh et al., 2006).
Thus, in order to have an additional proof that uridine residues are directly contacted by NANOS2, it would be interesting to analyse the crystal structure of NANOS2 zinc fingers in complex with an RNA containing the 'AUNAANU' motif. The crystal structure could also reveal whether uridine residues are contacted by the same hydrophobic amino acids found in *Drosophila*. However, this issue could also be addressed by using a different method, which would determine the exact NANOS2 amino acids bound to the RNA targets, by using UV-cross-linking coupled to mass spectrometry (Kramer et al., 2014; Shchepachev et al., 2019).

Further CRAC analyses revealed that the motif enrichment did not correlate with CRAC signal intensity, which is associated with NANOS2 targets. This suggested that NANOS2 motif could be necessary but not sufficient to determine NANOS2 binding intensity to RNA. Indeed, many factors can contribute to achieving target selectivity. For instance, the 3' UTR structure (see Chapter 1), can affect the accessibility of RBPs to certain sequences, thus controlling their binding along the RNA. I have shown that the NANOS2 motif present in *Taf7l* 3' UTR is predicted to be in a low energy region, which spans one or two loops. This preliminary finding suggests that NANOS2 might prefer binding to very accessible regions and that it might be a ssRNA binding protein, like other NANOS proteins (Hashimoto et al., 2010b). However, it would be interesting to assess at a global level whether NANOS2 targets adopt

a conformation similar to the one in *Taf7l* 3' UTR, in the proximity of the motif. This could be addressed in the future by analysing transcriptome-wide the secondary structure of 3' UTRs (Bevilacqua et al., 2016; Jacobs et al., 2012; Kwok et al., 2015), in SSC lines. This analysis could also reveal whether certain structural elements are more enriched in those targets with higher CRAC signal and thus explain how NANOS2 binding intensity is achieved *in vivo*.

Moreover, complex secondary structures could bring closer sequences which are normally far away. This could explain why not all NANOS2 targets had the motif adjacent to the direct binding site.

Importantly, the fact that NANOS2 binds to RNA with sequence specificity allows predicting NANOS2 targets also in other cells of interest.

The molecular mechanisms through which NANOS2 maintains mitotic quiescence and suppresses meiosis in male gonocytes is not fully understood. However, we could address this by sorting gonocytes from E13.5-15.5 and performing RNA-seq. Subsequently, we could determine which mRNAs contain the NANOS2 motif within the 3' UTRs. This might lead to the identification of direct NANOS2 targets in male gonocytes and help to gain insight into NANOS2 functions during early male germ cells development.

Additionally, NANOS2 has an overlapping expression with NANOS3, from E13.5 to E14.5, and it was also shown that they have partially redundant functions during early germ cell development (Suzuki et al., 2007). Their redundancy might be due to the presence of the same zinc fingers in both proteins. This could lead them to associate with a very similar pool of mRNAs and thus regulate analogous biological processes. To address this, we could also identify NANOS3 targets by CRAC. To circumvent the scarcity of male gonocytes *in vivo*, we could engineer *Nanos3*^{TAG} ESCs lines and then differentiate them into PGC-like cells *in vitro* (Hayashi et al., 2011).

Furthermore, NANOS3 is also present in the adult testis, where it has a broad expression, from the undifferentiated spermatogonia to the c-kit-differentiating

cells. Given the expression profile of NANOS2 and NANOS3 in relation to GFRα1 (Suzuki et al., 2009), it is likely that a small proportion of the GFRα1-positive Apaired and Aaligned(4) SSCs co-express both NANOS2 and NANOS3. Thus, we could assess whether they exert partially redundant functions in this fraction of the self-renewing SSCs pool. To address this, we could verify whether they share some of the mRNA targets in post-natal SSCs, by engineering a *Nanos3*^{TAG} mouse allele, deriving SSCs and by performing CRAC.

Moreover, *NANOS* genes have been associated with infertility as well as multiple cancers in human (De Keuckelaere et al., 2018). In order to investigate the molecular functions of the human NANOS proteins, it could be useful to predict NANOS targets, in both healthy and pathological human tissues. This could be addressed by analysing the transcriptomic profiles of human samples, and by looking for mRNAs which contain the 'AUNAANU' motif.

Since *NANOS1* was shown to be associated with human infertility (Kusz-Zamelczyk et al., 2013), it would also be interesting to investigate the molecular role of *NANOS1* during human germ cell development. This could be addressed by engineering *NANOS*^{TAG} human induced pluripotent stem cells (hiPSCs) lines, which would be differentiated into human primordial germ cell-like cells (hPGCLCs) *in vitro* (Sasaki et al., 2015), and used to perform CRAC.

3.9.3 Non-coding RNAs constitute a minor but novel class of targets bound by NANOS2 in SSC lines

mRNA is by far the most enriched class of transcripts bound by NANOS2, in SSC lines. Additionally, our data reveal for the first time that NANOS2 can also bind to non-coding RNAs (ncRNAs), although to a small extent.

Notably, miRNA was the most enriched class of ncRNAs bound by NANOS2 (~3% of all targets). Other RBPs, such as Lin28, were shown to associate with a tiny fraction of miRNAs, by performing CLIP in mouse ESCs (0.07% of all Lin28 targets) (Cho et al., 2012). However, the consequences of Lin28-miRNA interactions are known to be functionally relevant for development, from *C. elegans* to human cells (Tsialikas and Romer-Seibert, 2015). Similarly, part of the miRNAs bound by NANOS2 could have a critical role in SSCs maintenance, although they comprise only a small proportion of all its targets. To assess whether NANOS2 could regulate gene expression through miRNAs, we could verify the presence of *bona fide* miRNAs-mRNAs hybrids in SSC lines, by performing CLASH (Helwak and Tollervey, 2014), which is a variant of the CRAC. Moreover, by considering only the miRNAs with tissue-specific expression pattern as well as the ones evolutionary conserved, we could select a few candidate miRNAs for functional validation.

Overall, these analyses would allow clarifying whether there are indeed miRNA-mRNA in SSCs and whether these interactions have a biological relevance for mouse germ cell development.

Furthermore, CRAC analysis showed that many miRNAs bound by NANOS2 are located on the X-chromosome. Part of these are also transcribed from the same genomic region, and interestingly, this group of miRNAs has been recently shown to be testis-specific and highly conserved in mammals, and thus was named SpermiR (Zhang et al., 2019). The genetic deletion of single miRNAs from this locus did not cause major phenotypes in mice due to their possible redundancy. Thus, it would be interesting to delete the entire family of SpermiRs in mouse and to assess the consequences on spermatogenesis, as also suggested by Zhang and colleagues.

Moreover, SpermiRs are positioned between two highly conserved genes, *Slitrk2* and *Fmr1*, present both in mouse and human. *Fmr1* is the Fragile X mental retardation syndrome-related 1, and its loss was also associated with primary ovarian insufficiency (Wittenberger et al., 2007). Since also human NANOS3 was related to primary ovarian insufficiency (Santos et al., 2014; Wu

et al., 2013), it might exist a molecular relationship among hNANOS3, SpermiRs, and FMR1, in human ovaries.

I also reported that NANOS2 binds to another conserved group of miRNAs (miRNA25, miRNA106b, miRNA93), which overlaps to the Mcm7 locus and that is involved in cell proliferation and apoptosis (Mehlich et al., 2018). Thus, if we confirmed that these miRNAs formed hybrids with mRNAs in SSC lines by CLASH, we could also investigate their functional relevance for spermatogenesis, by deleting the locus from the mouse genome. Moreover, this locus is also involved in a variety of cancers in human (Mehlich et al., 2018). Given that also human NANOS proteins are dysregulated in cancer (De Keuckelaere et al., 2018), we could investigate the molecular interaction between these miRNAs and human NANOS, by doing CRAC from human cell lines. Overall, this could reveal novel mechanisms underlying tumour progression.

In the future, to gain insight into NANOS2 binding intensity to miRNAs, we could normalise CRAC signal with miRNAs abundance, by performing small-RNA-seq from SSC lines, as already done for the mRNA targets. This might change the ranking of the miRNA-CRAC list which was analysed so far and could highlight different miRNAs in the top of the CRAC list.

Overall, the fact that NANOS2 can bind to a small fraction of ncRNAs in mouse SSC lines, it might suggest that also other NANOS proteins could do so, in other cells and organisms. Thus, further investigations are needed to verify whether ncRNAs contribute to the intricate network of post-transcriptional gene regulation, mediated by NANOS proteins in germ cells.

3.9.4 NANOS2 binds to mRNAs involved in cellular metabolism and include mTORC regulators

In order to investigate the identity of the ~1400 transcripts identified by CRAC, we performed GO analysis. Interestingly, this revealed that NANOS2 targets are enriched in a variety of metabolic processes, including biosynthesis of nucleic acids, proteins, organic, and aromatic compounds. Moreover, protein modification, ubiquitination, translation, and chromatin organisation were also found to be moderately enriched in NANOS2 targets.

NANOS2 is generally thought to preserve SSCs self-renewal by repressing mRNAs involved in spermatogonia differentiation. However, GO analysis did not identify specific processes related to the differentiation of male germ cells, such as meiosis. NANOS2 was previously shown to bind to *Sohlh2*, *Dazl*, *Taf7l*, and *Dmrt1* mRNAs, by native RIP in SSC lines (Zhou et al., 2015). All these mRNAs are important for SSCs differentiation (Cheng et al., 2007; Kimmins et al., 2004; Pointud, 2003; Zhang and Zarkower, 2017). However, we could confirm only *Taf7l*, and *Dmrt1* as bona fide targets, by CRAC. This indicated that NANOS2 could directly repress these two differentiation-related targets in SSC lines, as previously hypothesised (Zhou et al., 2015).

It is possible that *Sohlh2* and *Dazl* were identified by RIP because of their abundance in SSC lines. Since native RIP does not stabilise RBP-RNA complexes before cell lysis, many post-lysis interactions can occur *in vitro* (Mili and Steitz, 2004), although they are not physiological. Therefore, although *Sohlh2* and *Dazl* mRNAs are important during later stages of spermatogenesis (Phillips et al., 2010; Suzuki et al., 2012b), my data suggests that most likely NANOS2 is not the direct link between these mRNAs and SSCs differentiation. Thus, *Dazl* may be directly bound by NANOS2 only in male gonocytes, as suggested by RIP (Kato et al., 2016), but not in post-natal SSCs. However, to address whether this interaction is indeed direct, we could perform NANOS2 CRAC from *in vitro* differentiated PGC-like cells (as previously explained).

Interestingly, many targets bound by NANOS2 with high intensity (top positions in CRAC list) have their highest expression in the central nervous system. Among others, there were Wdr89, Snrpb2, Fbxo45, Epc2, and Fbxo33 genes. Moreover, *Ddx3y*, the first CRAC target, is also involved in neuron development (Vakilian et al., 2015). NANOS1 in mouse is expressed in the central nervous system (Haraguchi et al., 2003). Similarly, dNANOS is also expressed in dendrites, where it regulates their morphogenesis (Ye et al., 2004). Additionally, dNANOS also regulates sodium current in motoneurons by repressing Para mRNA, in concert with PUMILIO (Muraro et al., 2008). Given the importance of post-transcriptional gene regulation in the brain (Kiebler et al., 2013), it is possible the RBP NANOS2, or its paralogs could have a critical role in this organ, also in mouse and not only in *Drosophila.* It would be interesting to explore NANOS2 localisation in the central nervous system by performing immunofluorescent staining of mouse brains. To do so, we could take advantage of the *Nanos2*^{TAG} allele, that has a GFP tag. Thus, by investigating NANOS expression in the brain, we might expand our knowledge on these proteins and potentially discover novel functions.

Overall, the GO analysis revealed clearly that NANOS2 targets are mainly involved in cellular metabolism. In general, metabolism is highly regulated during spermatogenesis, where each stage is adapted to particular metabolic needs, also with the support of Sertoli cells (Rato et al., 2012). This is in part due to the compartmentalised environment inside the seminiferous tubules, which partially limits the availability of metabolites, as, for instance, oxygen (Rato et al., 2012). Recently, single-cell RNA-seq of human SSCs (hSSCs) revealed that upon differentiation, hSSCs undergo four developmental states, one of which is characterised by a major metabolic transition (Guo et al., 2017). Additionally, human spermatogonia are known to rely on glucose to produce ATP, and this might be provided by blood components (Rato et al., 2012). Importantly, the regulation of glycolysis via Myc is essential for SSCs self-renewal in mouse (Kanatsu-Shinohara et al., 2016).

One of the most critical regulators of cellular metabolism is mTOR signaling. By sensing the availability of nutrients and growth factors, this pathway controls major processes such as glucose metabolism, lipids and nucleotides biosynthesis, but also protein translation. Overall, this allows mTOR coordinating cellular growth and proliferation (Saxton and Sabatini, 2017).

Different stem cells populations have been shown to balance their self-renewal and differentiation by controlling cellular growth via the mTORC pathway (Castilho et al., 2009; Gan and DePinho, 2009). Among other stem cells, also mouse SSCs were shown repress mTORC hyper-activation, in order to maintain their self-renewal ability. This is also achieved by regulating the expression of Redd1, a mTORC1 inhibitor, which is induced by the transcription factor PLZF (Hobbs et al., 2010). Importantly, the deletion of only one mTORC1 inhibitor, Tsc1, causes subfertility in mice. This phenotype is due to premature SSCs differentiation and germ cell depletion, which are caused by the hyper-activation of the mTORC1 pathway (Wang et al., 2016). Interestingly, NANOS2 was shown to control SSCs metabolism by repressing mTORC1 in SSC lines (Zhou et al., 2015). Notably, NANOS2 interacts with mTOR and confines it to stress granules, which results in general translation repression and inhibition of the mTORC1 pathway.

Although our GO analysis did not specifically detect the mTORC1 pathway, I could identify different positive regulators of the mTOR signaling, among the first 500 NANOS2 CRAC targets.

Therefore, we speculate that NANOS2 could post-transcriptionally control the expression of many metabolic genes, including five key mTORC regulators (Lamtor3, Rheb, Xbp1, Rictor, and Mios). Importantly, their potential repression mediated by NANOS2 could slow down metabolism and cellular growth in SSCs, in order to ultimately prevent the stem cell pool depletion over time.

3.9.5 One-third of NANOS2 targets takes part in the transition from self-renewing to transit-amplifying spermatogonia

After determining the identity of NANOS2 targets by GO, we sought to determine their expression profile throughout adult spermatogonia development. To this end, we sorted three populations of spermatogonia ex-vivo, from adult testes, which represented three different stages: the self-renewing SSCs pool, marked by Gfrα1, the transit-amplifying spermatogonia marked by Miwi2-Tom, and the differentiating spermatogonia, which express both Miwi2-Tom and c-kit. Then, we performed mRNA-seq from these three populations, and we applied the Markov clustering algorithm to define mRNA expression profiles. This analysis revealed that the majority of genes whose expression changes during the stages considered is characterised by a sharp upregulation from Gfrα1^{+} to Miwi2-Tom^{+} cells (Cluster1, ~6000 genes). Importantly, we discovered that approximately one-third of NANOS2 targets (469 transcripts) were also present in this cluster. Moreover, also another important group of genes (cluster 2, ~4000 genes) underwent significant changes throughout spermatogonia development. This was constituted by genes down-regulated from the self-renewing to the transit-amplifying population (expression profile opposite to cluster 1). Interestingly, NANOS2 targets were under-represented in this cluster. Thus, given the expression profiles of cluster 1 and 2, the over-representation of NANOS2 targets in cluster 1 is in accordance with their under-representation in cluster 2.

Overall, our analysis suggests that once Gfrα1^{+} SSCs (which are all Nanos2^{+}) start to undergo differentiation (Miwi2-Tom^{+} cells), one-third of NANOS2 targets is upregulated. Importantly, NANOS2 targets are enriched in processes such as cellular metabolism and include mTORC positive regulators. Therefore, we speculate that NANOS2 could repress cellular growth in Gfrα1^{+} spermatogonia. This mechanism would safe-guard SSCs quiescence, and protect them from excessive proliferation, thus inhibiting SSCs commitment towards differentiation.

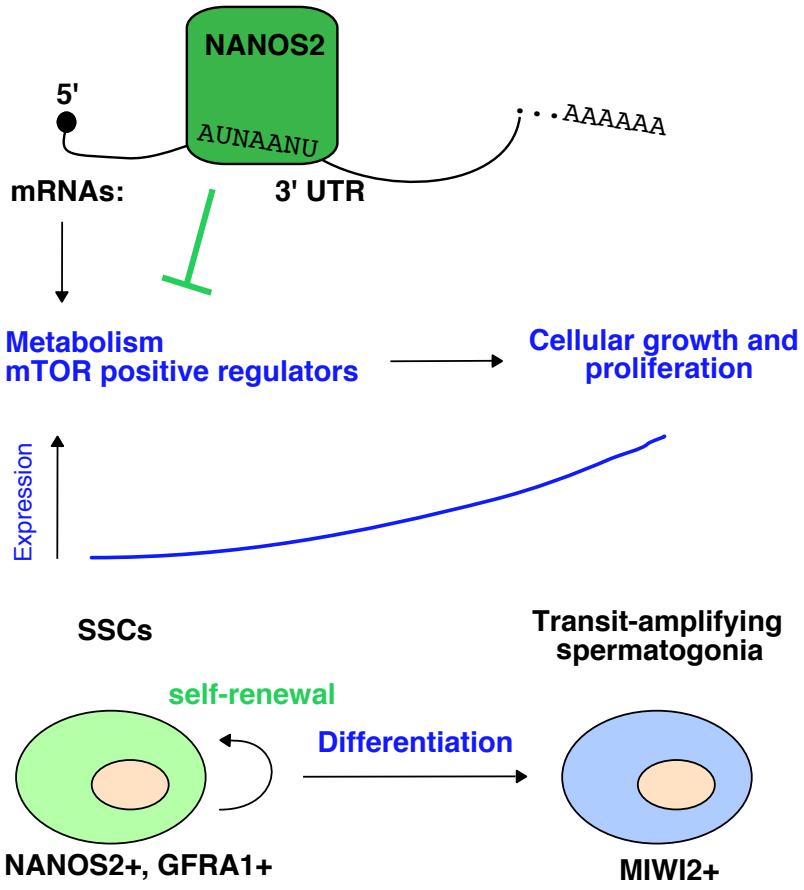


Figure 3.20. Possible working model for NANOS2 in SSCs.

CRAC showed that NANOS2 directly binds to the 3' UTR of mRNAs, enriched for a novel motif. These mRNAs are involved in metabolism, and they also include positive regulators of mTORC. Metabolic genes and the mTORC pathway are strictly related to cellular growth and proliferation (Ma and Blenis, 2009). One-third of NANOS2 targets are upregulated during the transition from self-renewing SSCs (GFrα1^{+}) to transit-amplifying spermatogonia (Miwi2^{+}). Thus, we speculate that NANOS2 could preserve SSCs self-renewal through the post-transcriptional repression of metabolic mRNAs, which would slow down metabolism, and promote the quiescence of GFrα1^{+} spermatogonia.

However, it is worth noting that NANOS2 targets were only slightly enriched in cluster 1 (P-value < 0.05) and moderately depleted in cluster 2 (P-value <0.01). This weak statistical significance might be due to a number of reasons. First of all, we compared two datasets generated from the high-throughput sequencing of cells isolated from two different sources. Indeed, CRAC was performed in SSC lines, and RNA-seq for Markov clustering from ex-vivo sorted spermatogonia. This discrepancy might implicate a certain degree of background noise in the comparative analysis. The extent of this variability could be addressed by comparing the transcriptional profile of SSC lines and Gfrα1-positive cells. To do so, we could perform a PCA analysis between the RNA-seq from SSC lines and the one from ex-vivo sorted Gfrα1^{+} cells.

Moreover, although all Gfrα1^{+} cells are also Nanos2^{+}, and ~70% of Nanos2^{+} cells express Gfrα1 (Suzuki et al., 2009), the two populations do not overlap completely. Since we isolated only Gfrα1^{+} cells, these SSCs do not take into account 30% of Nanos2^{+} spermatogonia. Thus, in order to comprehensively characterise the Nanos2^{+} population, in the future, we could employ a Nanos2-reporter allele, sort the Nanos2^{+} cells, and integrate our analysis with these additional data. This might reveal that more than one-third of NANOS2 targets participate in the transition from self-renewing to transit-amplifying spermatogonia.

However, it is also possible that genes in cluster 1 are not controlled exclusively by NANOS2. For instance, these genes could also be rapidly upregulated by NANOS2 targets, both at the transcriptional and post-transcriptional level.

Moreover, we cannot exclude the collaboration of more SSCs intrinsic factors present in Gfrα1^{+} cells, such as NANOS2, PLZF, and ID4. This might contribute to the massive changes in gene expression (~6000 genes) observed during the transition from self-renewing SSCs to transit-amplifying spermatogonia. Overall, this hypothetical cooperation could constitute a robust

mechanism to control the first step towards differentiation, during adult spermatogonia development.

Chapter 4. Characterisation of NANOS2 protein interactome and mRNA stability in SSCs

4.1 Introduction

In the previous chapter, I have explored NANOS2 RNA-binding transcriptome-wide in SSC lines by using CRAC. NANOS proteins mediate RNA-binding by the conserved zinc finger motifs within their C-terminus (Hashimoto et al., 2010b; Weidmann et al., 2016). In *Drosophila*, the genetic analysis of several *nanos* mutants lacking selective amino acids in the zinc fingers, proved that the zinc fingers were essential for all *NANOS* functions (Arrizabalaga and Lehmann, 1999). Moreover, NANOS2 zinc fingers were also shown to mediate protein-protein interactions by recruiting DND1 (Suzuki et al., 2016). Although the C-terminus is assumed to be necessary for NANOS2 function, it is not sufficient in mouse. Indeed, a NANOS2 mutant lacking the first ten amino acids in the N-terminus does not rescue defects in a *Nanos2*$^{-/-}$ mouse (Suzuki et al., 2012a). This N-terminal portion is essential for the direct interaction between NANOS2 and NOT1. The discovery of this important interaction raised the possibility that additional proteins could be recruited by NANOS2, through the same or different regions. Co-immunoprecipitation (co-IP) experiments have previously identified different NANOS2-interacting proteins, both in male gonads and in SSCs (Suzuki et al., 2010; Zhou et al., 2015). More recently, two proteomic approaches have been used to characterise NANOS2 interactome in embryonic male gonads (Suzuki et al., 2016; Zhou et al., 2017), revealing novel protein partners.

In this chapter, I have further investigated NANOS2 molecular function by determining its interactome in post-natal SSCs. To do so, we performed immunoprecipitation followed by mass spectrometry (IP-MS) from *Nanos2*$^{TAG/TAG}$ *in vitro* SSC lines.

Subsequently, we followed-up on the main molecular pathway revealed by the interactome, which was RNA decay. Therefore, I illustrate how we analysed transcripts stability in SSCs, by using a novel metabolic approach, called SLAM-seq (Herzog et al., 2017).

4.2 Endogenous NANOS2 protein interactome in SSC lines

To identify NANOS2 protein interactome in mouse SSCs, we performed immunoprecipitation followed by mass spectrometry (IP-MS), in collaboration with Dr. Much and Dr. Auchynnikava. For this purpose, we used the previously described *Nanos2*$^{TAG/TAG}$ SSC line, here called *Nanos2*^{TAG}, and, as a negative control, the untagged *Nanos2*^{WT} line, called *Nanos2*^{CTL}. A schematic in Figure 4.1 shows the entire procedure.

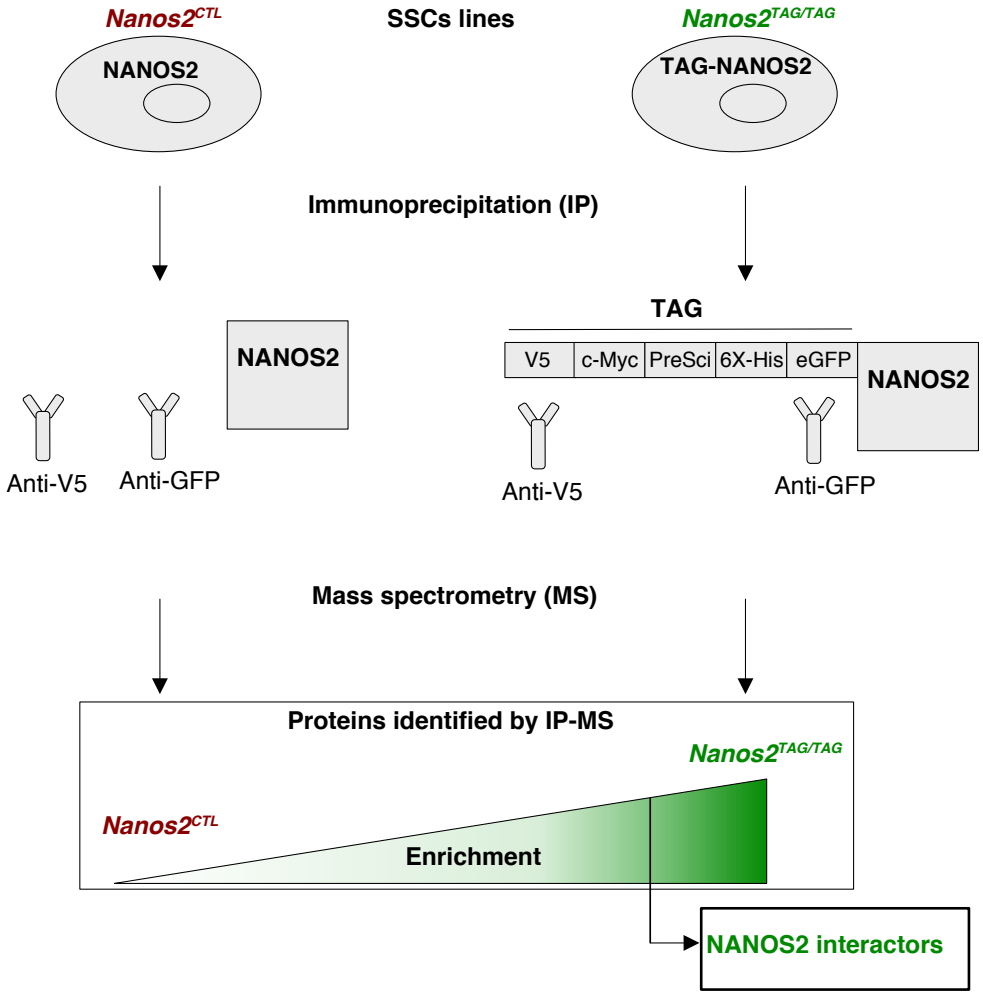


Figure 4.1. Schematic of the procedure used to determine NANOS2-protein interactome in SSC lines.

Two different SSC lines were used for Immunoprecipitation (IP) coupled to mass-spectrometry (MS). The first one (in red), *Nanos2*^{CTL}, expresses an untagged NANOS2 protein, and thus served as a control. The second SSCs line (in green), *Nanos2*$^{TAG/TAG}$, expresses TAG-NANOS2, whose tag contains the epitopes used for IP (V5 and eGFP). After immunoprecipitating NANOS2 from both SSC lines with either Anti-V5 or Anti-GFP antibodies, samples were subjected to MS. NANOS2 interactors (in green) were identified as proteins significantly enriched (four-fold enrichment and p-value < 0.05) in the *Nanos2*$^{TAG/TAG}$ IP, compared to the control IP (*Nanos2*^{CTL}). For more details see Figure 4.3.

SSCs cultures were expanded in 10 cm^2 plates, and each confluent plate was used for one experimental replicate. We prepared protein extracts by using a lysis buffer containing an intermediate salt concentration and non-ionic detergents, in order to preserve protein-protein interactions. In order to obtain more robust IP-MS results, we performed two immunoprecipitations in parallel, by taking advantage of the multiple tags in TAG-NANOS2. Thus, we incubated protein extracts with beads coupled to either Anti-V5 or Anti-GFP antibodies. Samples were washed with a mild buffer without salts. Half of the samples were also washed in the presence of RNase to assess whether protein-protein interactions depended on RNA. The remaining samples washed without RNase, also served as negative controls. Proteins were eluted from beads, and 20% of the eluate was used for quality controls. 10% of the eluate was separated by SDS-PAGE and transferred to a nitrocellulose membrane. First, to exclude an antibody's contamination, we stained the membrane with Ponceau (Figure 4.2 A). Importantly, we could not detect any band corresponding to the heavy and light chains of antibodies, suggesting that these were not co-purified in the eluates.

Then, to assess TAG-NANOS2 immunopurification, we probed the membrane with an anti-V5 antibody. As shown in Figure 4.2 A, NANOS2 was present in the input of *Nanos2*^{TAG} (second lane), and it ran at the expected molecular size (~47 kDa). A very faint band was detected only in the lane corresponding to the GFP-IP and not in the V5-IP, indicating that the protein was not efficiently immunoprecipitated (this was already observed when optimising IP for CRAC in SSCs). Another 10% of the eluate was used for SDS-PAGE and silver staining, which however did not show any clear NANOS2 band (Figure 4.2 B). Anyway, since only a small fraction of the IP was loaded into both gels (A and B), we assumed that a sufficient amount of protein could still be present in the remaining part of the IP (80%). Therefore, we continued with samples' preparation for MS.

After protein precipitation and digestion, processed samples were loaded onto the MS and subsequently analysed.

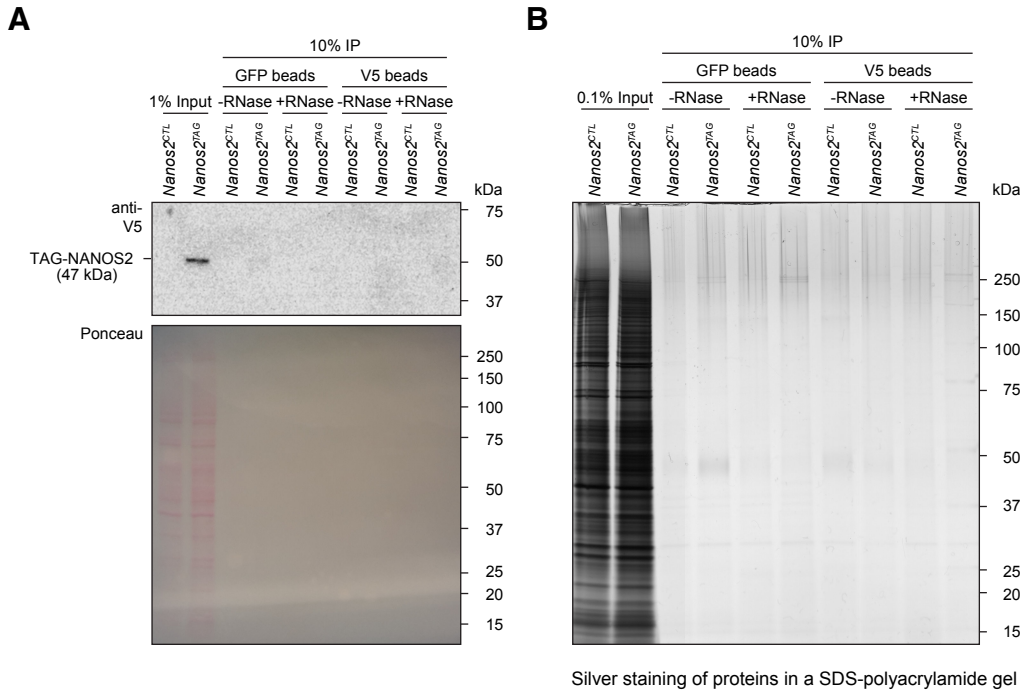


Figure 4.2. Endogenous NANOS2 immunoprecipitation (IP) and silver staining of protein gel.

A. Immunoblot probed with an Anti-V5 antibody showing two endogenous NANOS2 IP (with V5 and GFP beads), from *Nanos2*^{CTL} and *Nanos2*^{TAG} SSCs. The predicted molecular size of TAG-NANOS2 is ~47 kDa. Samples have been treated with (+RNase) or without RNase (-RNase). Ponceau staining of the same membrane is also shown below the blot. **B.** Silver staining of a protein gel loaded with 10% of the NANOS2 IP.

Proteins which were significantly enriched in the *Nanos2*^{TAG} IP compared to the control IP (*Nanos2*^{CTL}), were identified as NANOS2-interactors (see schematic in Figure 4.1). Notably, only interactors enriched with high confidence (four-fold enrichment and p-value < 0.05) were considered confident NANOS2-interactors and these are represented as labelled dots in the volcano plots (Figure 4.3). In these graphs, the x-axis represents the level of enrichment of precipitated protein, and the y-axis represents the level of confidence (calculated as described in the legend of Figure 4.3). Plots in A.

and B. represent MS results from anti-V5 IP and plots in (C) and (D) from anti-GFP IP. Results shown in (A) and (C) represents IP-samples not treated with RNase, whereas (B) and (D) were treated. Each graph shows results obtained from three replicates (N = 3), for each SSC line.

Importantly, NANOS2 was enriched in all conditions, although it was not previously detected by Western blot and silver staining in the IP. Overall, this suggested that NANOS2 was successfully immunopurified, for all the MS experiments.

Remarkably, a set of common interactors were found among the four MS. These included the CNOT1, 2 and 9 proteins, part of the CCR4-NOT deadenylase complex (CNOT), as well as TRIM21, an E3 ubiquitin protein ligase. Interestingly, these four proteins were found to interact with NANOS2 independently of RNA. Importantly, CNOT1 was the most enriched protein in all MS and experimental conditions tested. CNOT1 is the structural component of the CNOT complex, which also mediates protein-protein interactions with RBPs (see Chapter 1). Moreover, NANOS2-CNOT1 interaction was already described to occur in mouse male gonads by Co-IP, and it is also essential for Nanos2 function (Suzuki et al., 2010).

Therefore, our results show that the NANOS2-CNOT1 interaction is conserved also in post-natal SSCs, and that it could exert a critical role also in these cells. Moreover, almost all members of the CCR4-NOT complex were identified among the four MS. These included CNOT1, 2, 3, 9 and 10 as well as two catalytic subunits, CNOT7 and 8.

Additionally, TNKS1BP1, tankyrase1-binding protein, was detected in three out of four MS experiments. This protein is associated with TNKS1, Tankyrase1, which is a poly-ADP-ribosyl transferase involved in the post-translational modifications of proteins (Gibson and Kraus, 2012).

Collectively, our IP-MS results suggested that CNOT1, 2, 9, and TRIM21 proteins were the most robust NANOS2 interactors, as they were found with high confidence in all the MS performed. Given the essential role of the CNOT

deadenylase complex in RNA turnover (Collart and Panasenko, 2012), we speculated that NANOS2-CNOT interaction might have important implications for NANOS2 targets' stability in SSCs.

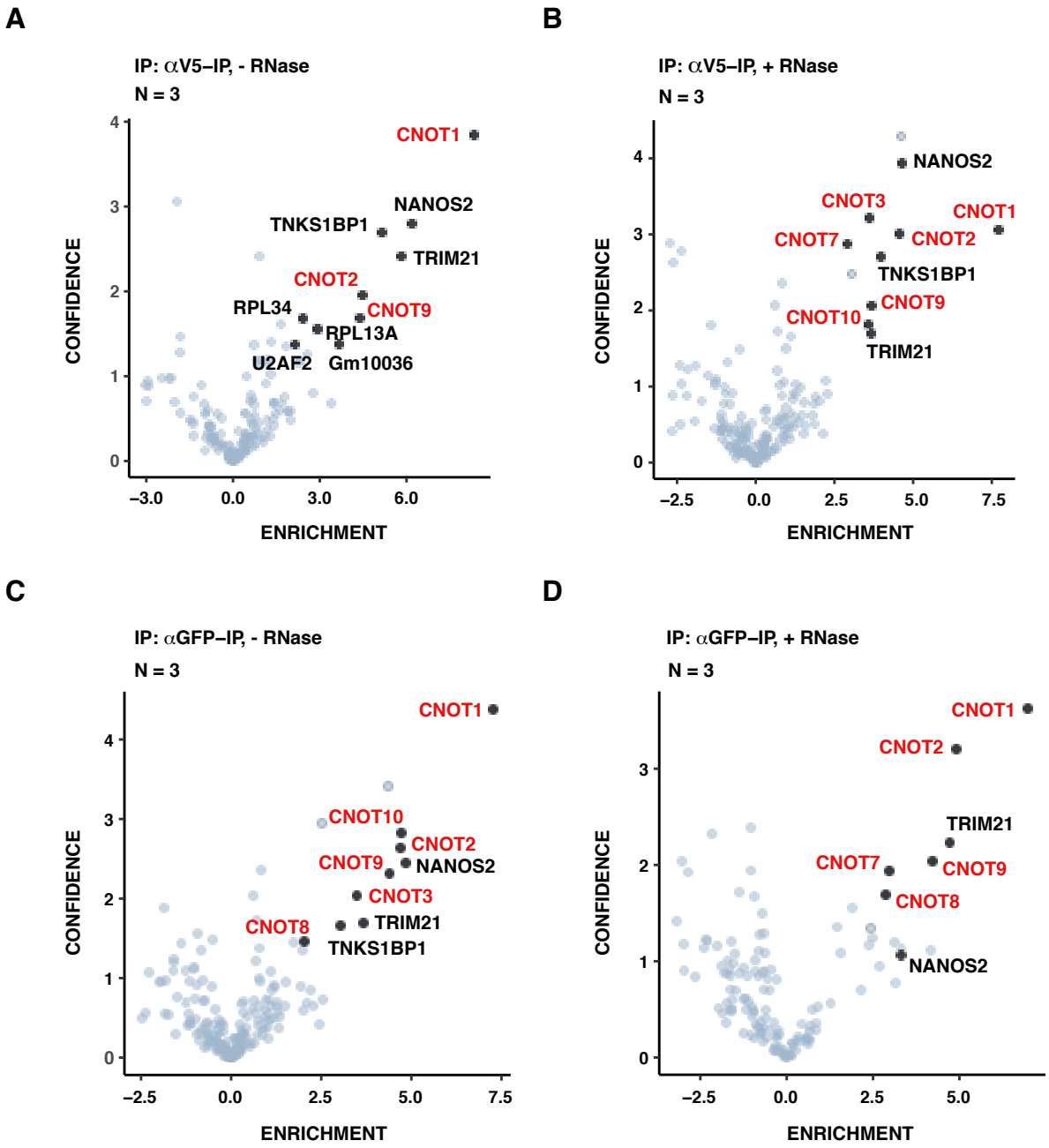


Figure 4.3. Volcano plots showing NANOS2-interactors identified by IP-MS, from SSC lines.

The endogenous NANOS2 was immunoprecipitated from *Nanos2*^{TAG} and from *Nanos2*^{CTL} (control) SSC lines. Immunoprecipitations (IP) were performed with Anti-V5 and Anti-GFP antibodies, and IP were treated with or without RNase. Plots in **A.** and **B.** display results from Anti-V5 IPs and plots in **C.** and **D.** show results from Anti-GFP IPs. IPs performed in **A.** and **C.** were not treated with RNase (- RNase), whereas in **B.** and **D.** RNase was added to the IP (+ RNase). The x-axis expresses the

enrichment of interactors in each IP-MS, defined as (log2(LFQ(Nanos2^{TAG})/LFQ(Nanos2^{WT}))), which is the logarithmised ratio of the average LFQ intensities for proteins detected in the *Nanos2^{TAG}*-IP over LFQ values of the same proteins in the *Nanos2^{WT}*-IP. The x-axis shows the confidence of the enrichment, defined as (-log10(p-value)), which is the negative logarithmised p-value obtained from a t-test. Proteins with at least four-fold enrichment and p-value < 0.05 are shown as labelled dots and are considered as interactors. NANOS2 is displayed in **D.** as a labelled dot, although it is enriched with slightly lower confidence than specified above. CNOT proteins are highlighted in red, as they are part of the same complex, (CCR4-NOT). Each graph shows results from three independent IP-MS replicates (N = 3).

4.3 Analysis of NANOS2 mRNA targets' stability by SLAM-seq in SSC lines

4.3.1 SLAM-seq: a metabolic approach to measure mRNA half-life

The characterisation of NANOS2 interactome in SSC lines showed that CNOT1 was the most enriched interactor, followed by other subunits of the CNOT deadenylase complex. Given the prominent role of the CNOT complex in RNA stability, we speculated that NANOS2-CNOT interaction might have important consequences on the half-life of NANOS2 RNA targets. In order to test this hypothesis, we sought to determine general mRNA stability in SSCs and to address whether NANOS2 mRNA targets were more unstable than the rest of mRNAs. To measure global mRNA stability in SSCs, I employed a pulse-chase labelling method, recently developed by Herzog and colleagues, called SLAM-seq (thiol(SH)-linked alkylation for the metabolic sequencing of RNA) (Herzog et al., 2017). This method allows tracing mRNA over time in cells and, importantly, it does not require high input of biological material, which made it suitable for *in vitro* SSC lines. Moreover, it allows transcriptome-

wide identification of mRNA stability, which is ideal in order to identify the half-life of NANOS2 CRAC targets globally.

Critical steps of the SLAM-seq method are shown in Figure 4.4 A and B. In brief, a nucleotide-analog, 4-thiouridine (s^4U), is supplied to cell media for a certain period of time, to allow its uptake by cells and its incorporation into newly synthetised RNAs. A critical step is to determine s^4U labelling time, which may vary across different cell types, depending on their specific RNA turnover rate. At the end of the labelling time, it is essential to stop s^4U incorporation into nascent transcripts, in order to allow discrimination between old (s^4U labelled) and new (unlabelled) RNA. This is achieved by supplying an excess of uridine to cell media, which is exchanged with s^4U. This results in the incorporation of uridine in place of s^4U into newly synthetised RNAs, from that moment onwards (t_0). RNA is isolated at different timepoints (t_x), in order to chase s^4U-containing RNA over time. After retrieving cells and extracting total RNA, s^4U is covalently modified (s^4U^*) through alkylation, by using iodoacetamide (IAA). IAA reacts with the thiol (S) group of s^4U, by nucleophilic substitution. This reaction converts s^4U to s^4U^*, which is much more stable than s^4U. It is worth to notice that the stabilization of s^4U by IAA-alkylation greatly enhances the efficiency of this method compared to other metabolic labelling protocols (Herzog et al., 2017). Since $s^4U^*$is chemically more similar to cytosine than uridine, s^4U^* is recognised as if it were a cytosine by the reverse transcriptase. Therefore, after reverse transcription (RT), guanosines are incorporated at s^4U^* positions. Adapters in RT primers allow subsequent PCR amplification of the cDNA. During PCR, guanosines are converted into cytosines, resulting in the thymidine-to-cytosine-conversion (T > C) of original s^4U. T > C conversions are further identified by high-throughput sequencing and used for the bioinformatic analyses of SLAM-seq data. Overall, these conversions allow discrimination between labelled (s^4U^*; T > C) and unlabelled (U; T) transcripts, and thus to retrace their fate over time. Therefore, transcripts which have high rate of T > C conversions in late time points are considered as the most stable ones.

In Figure 4.4 C, the expected RNA degradation kinetics are shown, after s^4U labelling. After reaching the steady-state level of s^4U incorporation, s^4U levels drop due to dilution with unlabelled uridine (green lines). By following the s^4U levels in mRNA over time, it is possible to obtain specific decay curves for each transcript (grey and black curves).

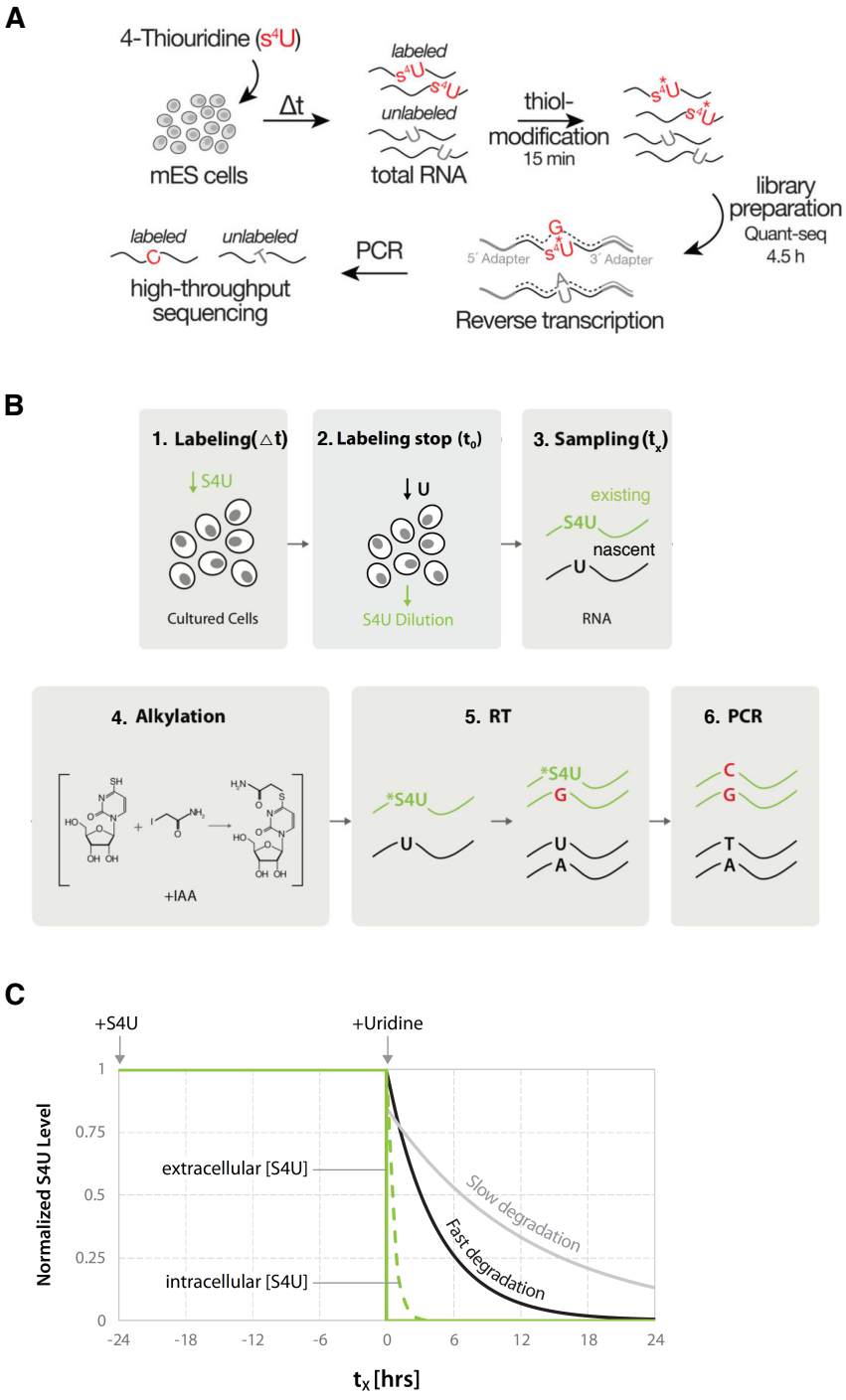


Figure 4.4. SLAM-seq work-flow and RNA degradation kinetics, adapted from (Herzog et al., 2017).

A. The diagram shows key steps of the SLAM-seq method. After adding s^4U, total RNA is isolated, and s^4U is stabilised by chemical modification. RNA is reverse transcribed using primers with adapters and guanosines are incorporated at s^4U positions. cDNA is amplified by PCR to generate libraries for sequencing. **B.** The

scheme highlights other critical steps of SLAM-seq. RNA is labelled with s^4U for a given time, Δt, until the reaction is stopped by diluting S4U with uridine, t_0. Cells are retrieved at different time points, t_x, and s^4U is stabilised by the addition of IAA (iodoacetamide), which modifies s^4U through an alkylation reaction (s^4U^*). Discrimination between existing and nascent RNA is achieved by detecting T>C mutations in sequencing reads. **C.** Kinetics of RNA degradation after s^4U labelling. Both slow and fast degradation kinetics can be detected by monitoring levels of s^4U labelled RNA over time (t_x).

4.3.2 SLAM-seq in SSC lines

In order to perform SLAM-seq in SSCs, first of all, I optimised the s^4U labelling in this cell type (see scheme in Figure 4.5 A). As previously explained, s^4U can exert cytotoxic effects, to the extent that may vary across different cell types. In order to minimise its cytotoxicity, I determined the optimal s^4U working concentration, which had a minimum impact on SSC viability, also called the inhibitory concentration, IC_{10}. In particular, this was defined as the s^4U concentration that would inhibit (kill) a maximum of 10% of SSCs, during a certain time (as described in (Herzog et al., 2017)). To do so, I tested a range of s^4U concentrations varying from 0 μM to 4000 μM in SSC lines. After culturing SSCs until they reached 70% confluency, I added s^4U to the cell media. Since SSCs are slow-dividing cells, we thought that they might also have a slow RNA turnover rate. Therefore, we opted for a relatively long labelling time (24 hours), and tested cell viability for twice the labelling period, for a total of 48 hours. SSC viability was assessed by counting cells' number at the longest time point (t_{48}). It is worth to notice that in the original SLAM-seq protocol, numbers of live cells are used for the cell viability assay, rather than just cell numbers. Although I tried to do so, this strategy did not allow me to determine the s^4U cytotoxic effect in SSCs successfully. In general, SSC death in culture occurs through cell detachment from colonies, so the only SSCs

suitable for experimental analysis are the ones which remain attached. When I analysed cell death, following s^4U incubation, the great majority of SSCs were found alive. Indeed, they displayed very low cell death rates (~1-4%), even at the maximum s^4U concentration (data not shown). Since cell death rates below 10% were not useful to determine s^4U working concentration, I concluded that measuring live cells was not the most appropriate way to detect s^4U toxic effect in SSC lines. Hence, I assumed that SSCs number would be more informative than live cells, and thus, I simply used this measure for the cell viability assay. After counting the number of SSCs retrieved from s^4U-treated samples, I compared them to SSCs number from the untreated ones (negative control). Their relative percentage was plotted over s^4U concentrations, as shown in the graph in Figure 4.5 B. Two technical replicates were performed for the cell viability assay (rep1 and rep2), using the *Nanos2*^{TAG} SSC line. Average values between the two replicates (red curve) were used to calculate IC_{10}. This analysis revealed that the optimal s^4U concentration that killed a maximum of 10% of SSCs, after a labelling time of 24 hours, was 100 μM.

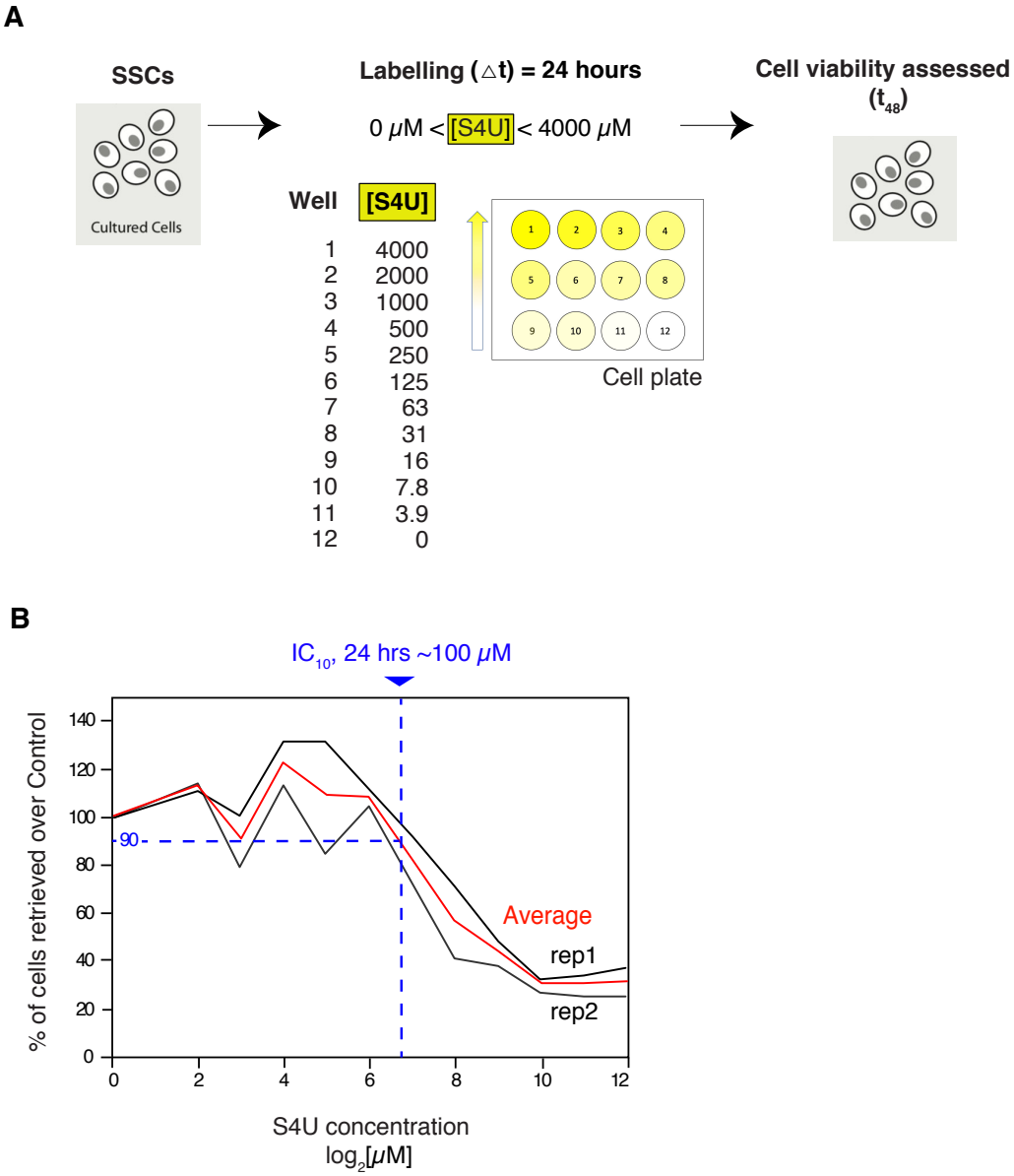


Figure 4.5. Cell viability assay in SSC lines used to determine s^{4}U cytotoxicity. **A.** A range of s^{4}U concentrations, [s4U], were tested on SSCs for a labelling time of 24 hours. After an additional 24 hours (t_{48}), SSCs were retrieved and counted to assess cell viability. **B.** The graph shows the numbers of SSCs retrieved when cultured with s^{4}U. The x-axis indicates [s^{4}U] used. The y-axis indicates the percentage of s^{4}U-treated SSCs over the untreated ones. The optimal working concentration of s^{4}U (IC_{10}) is shown in blue (100 μM). Two replicates (rep1 and rep2) were used for this assay, and their average values (curve in red) were used to calculate the IC_{10}.

After determining the s^4U working concentration, I performed the SLAM-seq experiment by using the *Nanos2*$^{TAG/+}$ and *Nanos2*$^{TAG/TAG}$ SSC lines. The experimental SLAM-seq strategy for SSCs is shown in Figure 4.6.

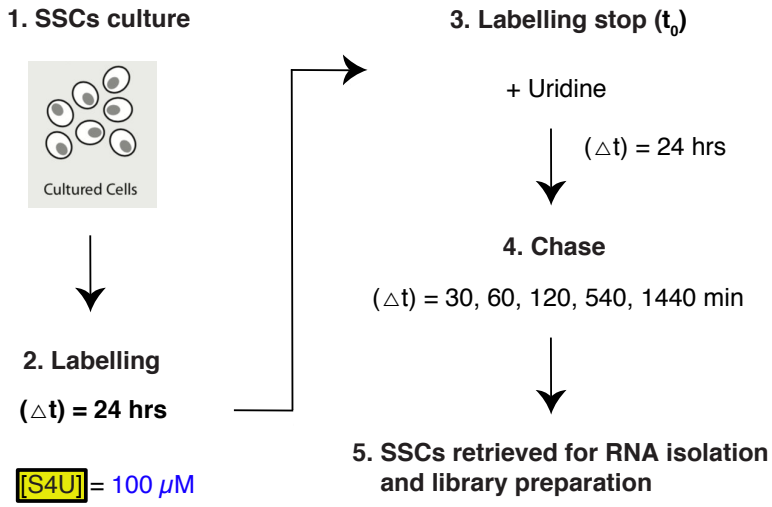
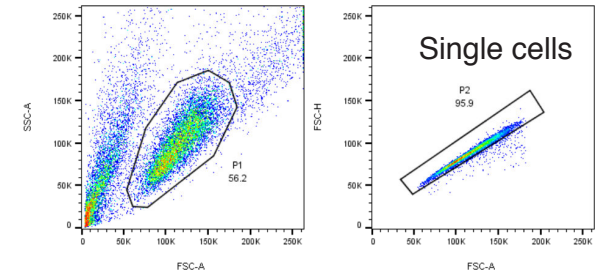


Figure 4.6. SLAM-seq work-flow in SSC lines.
RNA in SSCs was labelled with [s^4U] = 100 μM for 24 hours until the addition of unlabelled uridine stopped its uptake (t_0). To monitor RNA decay through s^4U chase, samples were collected at different time points (t = 30, 60, 120, 540, 1440 minutes). RNA was isolated, and libraries were prepared for sequencing.

In order to obtain sufficient material for RNA extraction, I cultured SSCs on 10 cm^2 plates and added s^4U in the cell media to a final concentration of 100 μM. I incubated SSCs with s^4U for 24 hours and stopped the labelling by adding an excess of unlabelled uridine (t_0). To follow RNA decay over time, SSCs were isolated at different time points (t = 30, 60, 120, 540, 1440 minutes), with the longest one set at 24 hours (or 1440 minutes). Each 10 cm^2 plate represented

an experimental time point, for a specific SSC line. SSCs were purified from MEFs co-culture by doing MEF depletion (see Chapter 2) and FACS. SSCs sorting strategy in shown in Figure 4.7 A. SSCs were first selected according to cell shape and size (FSC and SSC) and only living cells (negative to the DRAQ7 dye) were isolated. Approximately one million of SSCs was sorted for each experimental time point. Total RNA was extracted, and s^4U stabilised by IAA treatment (as previously explained). Purified SLAM-seq libraries had an average peak size of ~250 bp, as shown by the representative electropherogram in Figure 4.7 B.

A



Tube: T 2hr 5

Population	#Events	%Parent	%Total
All Events	18,742	####	100.0
P1	10,538	56.2	56.2
P2	10,101	95.9	53.9
P3	9,988	98.9	53.3
P4	9,872	98.8	52.7

B

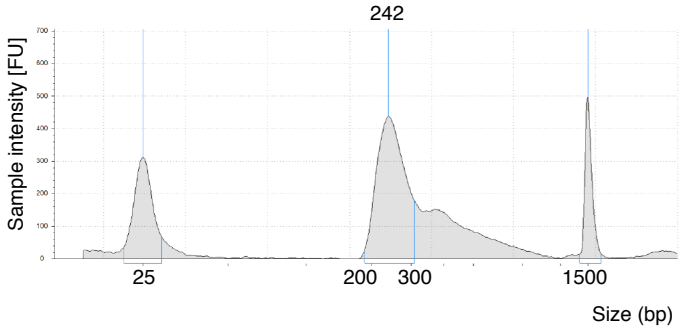


Figure 4.7. SLAM-seq in SSC lines: FACS strategy used to purify SSCs from MEFs and mRNA library.

A. Gating strategy used to purify SSCs from MEFs co-cultures. SSCs were sorted according to cell size (FSC-A, SSC-A) (P1). Single SSCs were isolated according to FSC-H, SSC-A, and SSC-H, FSC-A, respectively (P2, P3). Live SSCs were sorted by selecting the Draq7 negative cells (P4). The table shows numbers of SSCs sorted

(events) for each population (P1, P2, P3, P4) and their percentage with respect to the previous population (%Parent) and the original one (%Total). **B.** Example of an electropherogram, showing a typical SLAM-seq library obtained after cDNA amplification by PCR. The peak size of the mRNA library is ~ 250 bp.

Libraries were subjected to high-throughput sequencing, and the subsequent bioinformatic analysis was performed by Dr. van de Lagemaat.
By counting the T to C conversion rate within reads (see previous explanation), we determined the original s^4U levels in transcripts and thus their stability.
First of all, we assessed the degree of correlation between the two biological replicates (rep1, rep2). By analysing half-life values of individual genes in a scatter plot (Figure 4.8 A), we concluded that the two samples correlated very well with each other, as shown by the red line in the graph. Afterwards, we calculated global mRNA half-life in SSCs by using the T to C conversion rate from each experimental time point (Figure 4.8 B). When half-life for a given gene failed to be calculated in one of the two replicates, we excluded that gene from the analysis. Eventually, mRNA stability was determined for a total of 5806 genes. Overall, this analysis revealed that the average mRNA stability in SSCs was ~3.73 hours (or 224.5 minutes). Importantly, this value was extremely similar between the two replicates (224 and 225 minutes). Interestingly, although SSCs are slow-dividing cells, they displayed even a slightly faster RNA turnover than ESCs, which have an average mRNA half-life of ~3.9 hours (Herzog et al., 2017).

Additionally, we examined the decay kinetics of few individual transcripts, *Taf7l*, *Sycp1*, *Dmrt1*, and *Zfand6*, previously identified by CRAC (Figure 4.8 C). These showed different degradation dynamics, with *Sycp1* having the shortest half-life, and *Taf7l* the longest one.

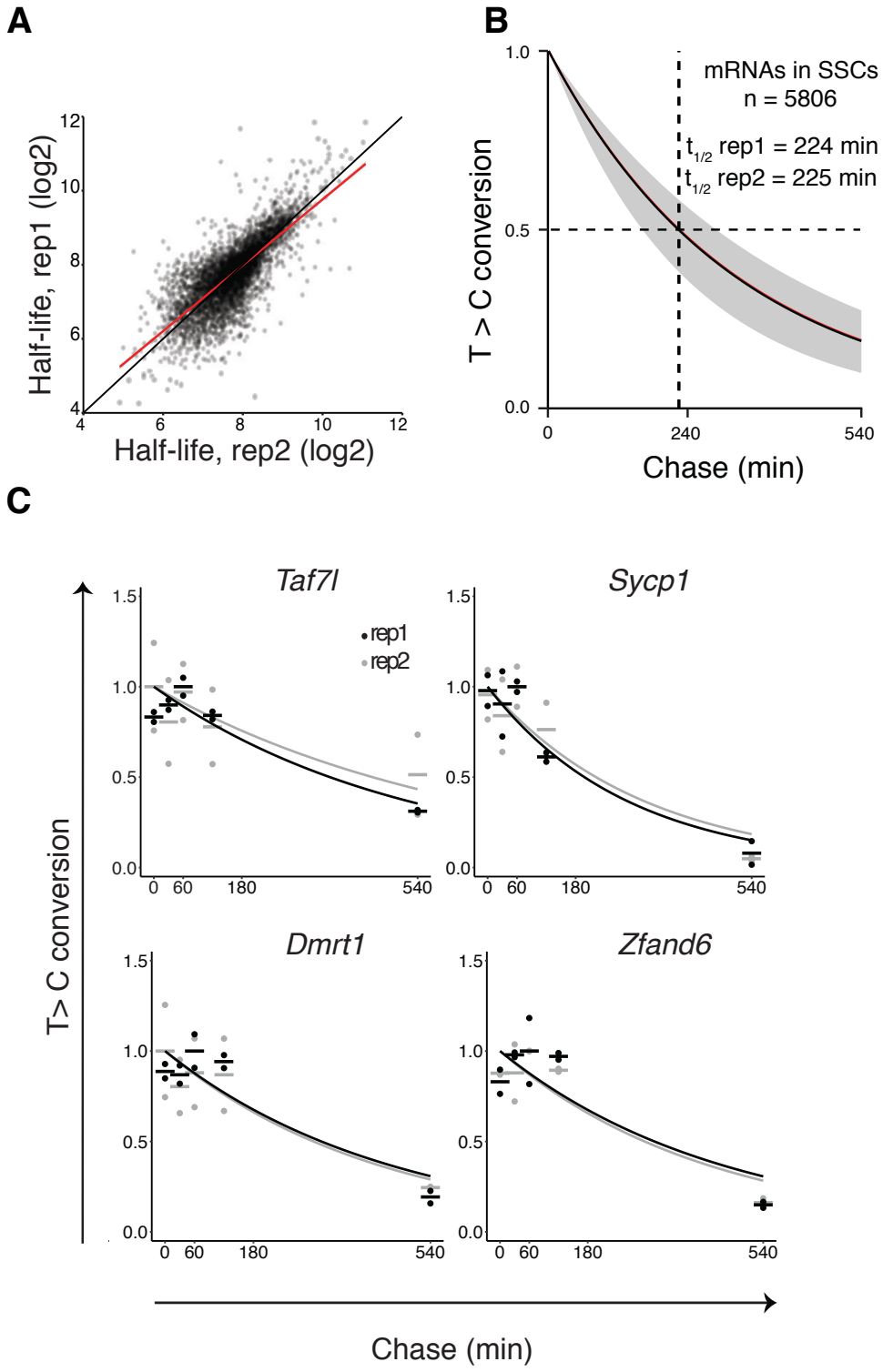


Figure 4.8. Analysis of mRNA stability in SSC lines measured by SLAM-seq.

A. In the scatter plot shown, each dot represents the half-life of an mRNA. The graph shows the correlation of half-life values expressed in minutes (logarithmic scale), between two biological replicates ('rep1' is *Nanos2*$^{TAG/+}$, and 'rep2' is *Nanos2*$^{TAG/TAG}$

SSC line). **B.** The graph shows global mRNA stability in SSCs. RNA half-life ($t_{1/2}$) was calculated by counting the T to C conversion rate in sequencing reads (y-axis), over time (minutes, on the x-axis). Half-life was determined for 5806 mRNAs, from two replicates (rep1 and rep2). **C.** Example of RNA decay curves for four distinct genes (*Taf7l, Sycp1, Dmrt1, Zfand6*), which are also NANOS2 CRAC targets.

However, in order to understand whether NANOS2 binding could affect mRNA stability at a global level, we analysed the cumulative distribution of NANOS2 targets' half-life. These targets were previously ranked according to CRAC signal intensity, normalised with mRNA abundance (see Chapter 3). Importantly, this list included 1428 genes, which we subdivided into two categories, depending on CRAC intensity. The first group included the top 500 genes, which had the highest CRAC signal (Figure 4.9, 'high' targets, in red). The second category of genes included the remaining targets, with lower CRAC signal, as well as non-targets ('other', in black).
First of all, transcripts with low or no CRAC signal ('other') had similar stability to the global one in SSCs, (~224 minutes). However, targets strongly bound by NANOS2 ('high') displayed a significantly shorter half-life, with an average of ~180 minutes (~3 hours), which is 44 minutes less than the global stability in SSCs. Importantly, these data were consistent between replicates (rep1, p-value < 10^{-15}; rep2, p-value < 10^{-10}). Furthermore, these data were really indicative of the majority of NANOS2 CRAC targets, since they were based on half-life values calculated for 74% of the 'high' targets (370 genes out of 500). Additionally, CRAC signal intensity positively correlated with transcript instability. Indeed, when smaller subsets of NANOS2 'high' targets were analysed (top 400, 300, 200), these revealed a continuous and increasing mRNA instability (data not shown).

Overall, these data indicated that the direct interaction between NANOS2 and mRNAs resulted in a significant reduction of their half-life in SSC lines.

Additionally, this effect increased for targets with very high CRAC signal. Importantly, these data constitute the first molecular proof that NANOS2 binding causes mRNA instability at the transcriptomic level, and in a physiological context.

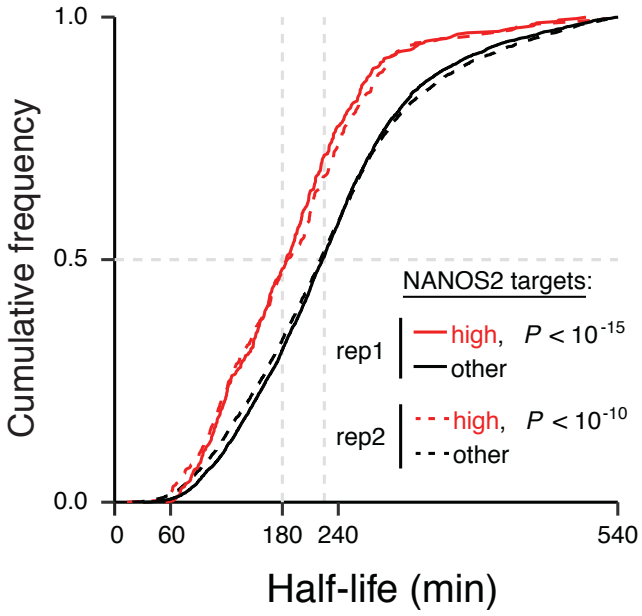


Figure 4.9. Analysis of NANOS2 targets stability in SSC lines.
S-curves show cumulative frequency of mRNA half-life for two distinct categories of transcripts. NANOS2 'high' targets represent the top 500 mRNAs from the CRAC list and are shown in red. NANOS2 'other' targets groups mRNAs with low CRAC signal and not bound by NANOS2, (in black). Results from two replicates are shown (rep1, rep2). P-values were calculated by using the Mann-Whitney *U* test.

So far, I characterised global mRNA stability for NANOS2 targets in SSCs and I also showed the correlation between CRAC signal and short half-life.
Subsequently, I investigated the stability of *Taf7l* and *Dmrt1* mRNAs, two targets identified both by CRAC and RIP (Zhou et al., 2015). Importantly, they are involved in SSCs differentiation, and it is believed that they are repressed

by NANOS2 in SSCs (see Chapters 1 and 3). Interestingly, both of them displayed remarkable mRNA stability. In particular, *Taf7l* had a half-life of 403.5 minutes and *Dmrt1* of 310 minutes, which are much higher values than the average half-life in SSCs. These data indicated that, although NANOS2 directly binds to these transcripts, its binding does not seem to trigger RNA instability.

I previously showed that NANOS2 could directly bind to mTOR regulators (Chapter 3), and potentially regulate the fate of these transcripts. Therefore, I evaluated whether also these mRNAs had a short half-life, similar to other NANOS2 targets. Interestingly, among the positive mTOR regulators which had a high CRAC signal (first 500 targets), I found that four out of five of them were more unstable compared to the average. These were represented by *Lamtor3*, *Rheb*, *Xbp1*, and *Rictor*, which had a half-life of 118.5, 197, 69.5, 122.5 minutes respectively (Figure 4.10). It is worth noting that *Xbp1* displayed a half-life of only 69 minutes; this value is significantly lower than the average for NANOS2 CRAC targets (~180 minutes) and even lower than the half-life of the very top targets (~100 minutes).
Overall, these data indicated that upon binding to specific mTOR regulators, NANOS2 triggered their decay. This mechanism might constitute a novel way through which NANOS2 represses the mTORC1 pathway in SSCs, in order to control cell growth and to ultimately preserve SSC self-renewal.

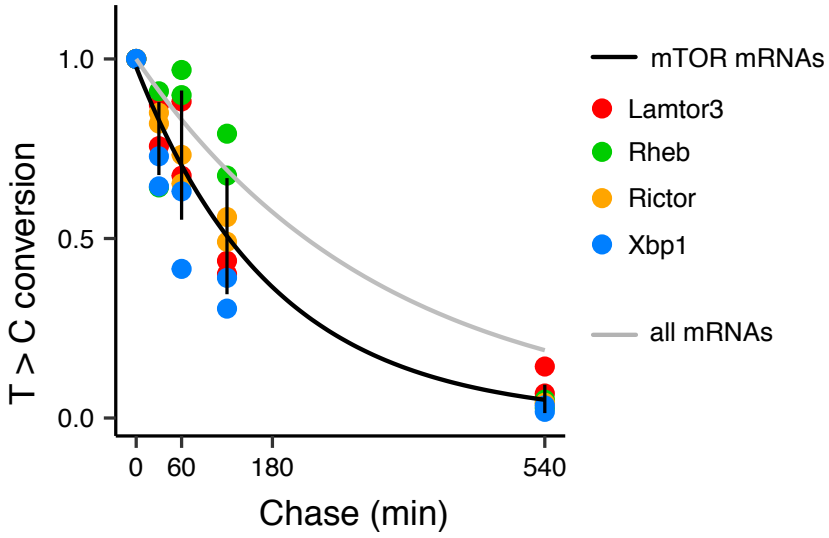


Figure 4.10. Decay curves representing four mTORC positive regulators (black) and the rest of mRNAs (grey).

The decay curve in black represents the combined half-life of the four mTORC positive regulators (*Lamtor3*, *Rheb*, *Rictor*, and *Xbp1* mRNAs). Black vertical lines are the standard deviations, for each timepoint. The decay curve in grey represents the combined half-life of all mRNAs analysed in SLAM-seq. Values from two biological replicates are reported in the graph.

4.4 Motif analysis in transcripts ranked according to their half-life

In the previous chapter, I have described that NANOS2 binds to mRNAs, which are specifically enriched in the novel 'AUNAANU' motif. In this chapter, I showed that NANOS2-mRNA targets have a short half-life. In order to further investigate the relation between transcript instability and NANOS2-RNA binding in SSCs, we sought to understand whether unstable transcripts also bore NANOS2 motif. To this end, we performed a motif enrichment analysis using the 3' UTR of mRNAs, whose half-life was determined by SLAM-seq.

We decided to use the Sylamer software (van Dongen et al., 2008), which looks for the presence of motifs in a given list of genes. These genes are sorted

according to a certain parameter of interest, which in our case, was half-life. The Sylamer bioinformatic analysis was performed by Dr. van der Lagemaat. It is worth to notice that we did not provide the software with the NANOS2 motif sequence, as we wanted to assess its presence in an unbiased way. As input for the analysis, we used the 3' UTRs of mRNAs ordered according to increasing values of half-life, previously calculated from two biological replicates (*Nanos2*$^{TAG/+}$ and *Nanos2*$^{TAG/TAG}$). Results of the Sylamer analysis are shown in Figure 4.11. In the graphs, the x-axis represents the 3' UTR of mRNAs, sorted according to half-life, and the y-axis shows the significance of each motif found. Dotted lines represent E-values thresholds. Coloured lines represent the landscape of enriched motifs across the ordered list of 3' UTRs. In particular, each coloured line is associated with a different motif, whose sequence is shown on the right. Grey lines represent motifs not linked to those 3' UTRs.

The first graph (Figure 4.11 A), shows results from the *Nanos2*$^{TAG/+}$ SSC line. Interestingly, Sylamer analysis identified three motifs as being moderately enriched in the 3' UTRs provided. Surprisingly, two motifs (the light blue, 'AUGAAAU', and the blue one 'AUGAAUU') were also compatible with the NANOS2 motif ('AUNAANU'). Importantly, these two motifs were enriched in the 3' UTRs having a short half-life (left part of the x-axis).

Figure 4.11 B shows results from the *Nanos2*$^{TAG/TAG}$ SSC line. Three motifs were found to be enriched in this analysis and, importantly, one of them matched the NANOS2 motif (in light blue, 'AUGAAUU'). Similar to the other replicate, also this motif seemed to be moderately enriched in transcripts with a short half-life. Interestingly, the 'AUGAAUU' motif, was already observed among NANOS2 targets, such as in *Otud4* mRNA (Chapter 3, Figure 3.14 B). Additionally, a different motif displayed a steep slope at the beginning of the x-axis, indicating an enrichment in 3' UTRs with an extremely short half-life (the one in purple, 'ACGCACG'). On the contrary, the third motif had a steep slope towards the end of the graph, indicating its enrichment in RNAs with a very long half-life (the blue one, 'UAAAUAU').

In summary, we identified NANOS2 motif in the 3' UTR of transcripts which displayed short half-life, from two biological replicates.
Overall, these data suggested that NANOS2 induced mRNA instability in SSCs by virtue of its binding to 3' UTRs, enriched for the ‘AUNAANU’ motif.

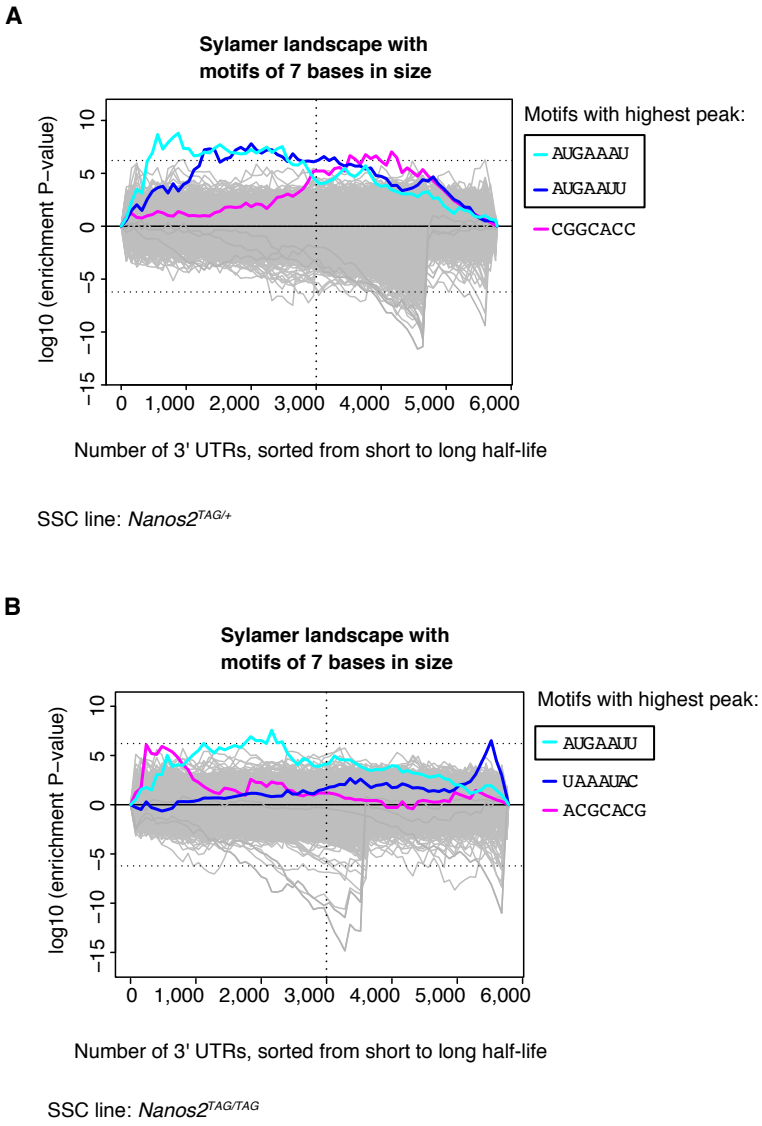


Figure 4.11. Motif enrichment analysis in 3' UTRs of transcripts ranked according to their half-life.

Sylamer plots show the landscape of motifs enriched in the 3' UTRs of mRNAs, which are sorted according to their half-life. 6,000 3' UTRs are displayed on the x-axis. The vertical dotted line in each graph is positioned in the middle of the x-axis (at 3,000 3' UTRs). The analysis was performed from *Nanos2*$^{TAG/+}$ (**A.**) and *Nanos2*$^{TAG/TAG}$ SSC lines (**B.**) Horizontal dotted lines are E-values threshold of 0.01. Each coloured line represents an enriched motif. The significance of the enrichment was calculated by using hypergeometric P-values. The sequence of the enriched motifs is shown on the right. Black rectangular boxes highlight sequences compatible with NANOS2 motif ('AUNAANU').

4.5 Discussion

4.5.1 NANOS2 mainly interacts with the CCR4-NOT deadenylase complex, in mouse SSC lines

By using the previously characterised *Nanos2*^{CTL} and *Nanos2*$^{TAG/TAG}$ SSC lines (Chapter 3), we could immunoprecipitate (IP) NANOS2 and identify the interacting proteins by mass spectrometry (MS). Thus, we determined the first NANOS2 protein interactome in SSC lines, which revealed some known partners as well as novel ones.

We discovered that members of the CCR4-NOT (CNOT) complex, notably CNOT1, 2 and 9, as well as TRIM21, constituted the most robust and confident interactors, as they were identified in all IP-MS performed. Moreover, their interactions with NANOS2 was independent of RNA, suggesting that they might interact directly in SSC lines.

Interestingly, CNOT1 was, by far, the protein enriched with the highest confidence. CNOT1-NANOS2 interaction was previously detected in male gonadal extracts (Suzuki et al., 2010, 2012a). Additionally, by using recombinant NANOS2 produced in bacteria, NANOS2 was shown to interact directly with CNOT1, through the NIM region (N-terminus, first 10 aa), which contacts the C-terminal part of CNOT1 (Suzuki et al., 2012a). CNOT1 was also shown to be the strongest interactors by performing IP-MS from male gonads (Suzuki et al., 2016). Moreover, the CNOT1, 3, 6L, 7, and 9 subunits were co-immunoprecipitated (co-IP) by NANOS2 from male gonads (Suzuki et al., 2010) and CNOT9 was co-IP by NANOS2 also from SSC lines (Zhou et al., 2015).

Thus, except for CNOT6L, we confirmed all these CNOT-NANOS2 interactions by IP-MS in mouse SSC lines, and we also found that additional subunits, such as CNOT2, 8, and 10 were associated with NANOS2.

Importantly, we determined that NANOS2 interacts with two out of four catalytic subunits of the CNOT complex, which are CNOT7 and CNOT8, in

SSC lines. Although CNOT6L was previously found to interact with NANOS2 both in gonads and post-natal testis (Suzuki et al., 2010), I could not identify it in our IP-MS. Thus, our data suggest that the CNOT6/6L might not be part of the NANOS2-CNOT complex in SSC lines. This would be compatible with the heterogeneity of the CNOT complex, which is thought to be composed of both integral and optional subunits (Lau et al., 2009; Winkler and Balacco, 2013).

Moreover, *Cnot6-6l* mRNAs are only moderately expressed in SSCs (Gfra1$^+$ Nanos2$^+$ cells) but are highly upregulated during the transition from undifferentiated (Miwi2$^+$ cells) to differentiating spermatogonia (c-kit$^+$ cells) (from our mRNA-seq data). On the contrary, *Cnot7* mRNA has an opposite expression pattern of *Cnot6/6l* and the same expression profile of *Nanos2*. Thus, these data suggest that the composition of the CNOT complex might change during adult spermatogonia development.

Moreover, *Cnot6* mRNA is expressed at a very low level in SSC lines (our mRNA-seq data, ~7 fpkm). Interestingly, *Cnot6* mRNA is also one of the strongest NANOS2 targets (our CRAC data, ranking position ~100) and, in sea urchin, *Cnot6* mRNA is also found amongst NANOS targets (Swartz et al., 2014). Therefore, these findings suggest that NANOS2 could regulate the dosage of some *Cnot* mRNAs (e.g., *Cnot6*) at the post-transcriptional level, thereby controlling the composition of the CNOT complex in SSCs. However, to confirm that the levels of the *Cnot* transcripts reflect their protein expression levels, we should perform western blotting using specific antibodies against the CNOT6/6L and CNOT7/8 subunits, in SSC lines.

Moreover, it is known that some RBPs preferentially bind to certain CNOT catalytic subunits rather than others. For instance, ZFP36L2 preferentially interacts with CNOT6L, rather than CNOT7, in Hela cells (Sha et al., 2018). Similarly, NANOS2 could prefer to bind to CNOT7 compared to CNOT6/6L, even if these subunits were equally expressed at the protein level, in SSC lines.

Furthermore, in yeast and human cells, CAF1 (CNOT7/8) takes part in one of the two specific phases of the deadenylation process, since it preferentially

shortens poly(A) tails not covered by PABP (Bresson and Tollervey, 2018) . Thus, NANOS2 in complex with CNOT7/8 could also be associated with this particular deadenylation stage, if this process really occurred on NANOS2 targets.

We also identified with high confidence two novel NANOS2 interactors. As previously mentioned, the E3 ubiquitin ligase TRIM21 was bound by NANOS2 in each condition tested, also independently of RNA. Since it was already shown that another E3 ubiquitin ligase, NEDD4, ubiquitylates NANOS2 in post-natal SSCs (Zhou et al., 2017), it is possible that also TRIM21 could do so, in order to modulate NANOS2 expression level or RNA-binding ability. To assess this, we could test whether NANOS2 directly interacts with TRIM21, by performing co-IP experiments in bacteria, after overexpressing a GST-NANOS2 and a tagged TRIM21. Moreover, it could be interesting to assess whether NANOS2 is ubiquitylated by TRIM21 by using an *in vitro* ubiquitylation assay.
Alternatively, it is also possible that TRIM21 directly interacts with CNOT to modulate the activity of the complex. For instance, it was shown that CNOT7 can be ubiquitylated by the E3 ubiquitin ligase, MEX-3C, and this results in the regulation of its deadenylase activity, in human cell lines (Cano et al., 2015a). Moreover, since we did not detect the E3 ubiquitin ligase CNOT4 in our IP-MS, it is also possible that TRIM21 could replace CNOT4 in the NANOS2-CNOT complexes, in SSC lines.

Additionally, also the tankyrase1-binding protein (TNKS1BP1) was identified as one of the most confident NANOS2 interactors. Interestingly, this protein was previously shown to associate with CNOT (Wahle and Winkler, 2013), although its role in the complex is not known. However, this suggests that TNKS1BP1 might not associate directly with NANOS2, but that it is part of the CNOT complex in SSC lines. TNKS1 mediates protein parsylation through its poly-ADP-ribosyl transferase domain (Gibson and Kraus, 2012), and this

activity is involved in a variety of cellular processes, which include DNA repair, transcription, regulation of ubiquitylation (Gibson and Kraus, 2012), RNA metabolism, and modulation of RNA stability (Ke et al., 2019). Interestingly, the human TNKS1 seems to have its highest expression in the adult testicular tissues (NCBI database), indicating its potential involvement in spermatogenesis. Thus, it could be interesting to investigate the relevance of TNKSBP1 during spermatogenesis, by deleting the Tnks1bp1 gene in mouse. Moreover, we could also assess whether components of the CNOT complex or NANOS2, are subjected to parsylation in SSC lines.
Overall, both parsylation and ubiquitylation have been related to RNA metabolism and, in particular, to RNA decay (Cano et al., 2010, 2015b; Hildebrandt et al., 2017; Ke et al., 2019). Therefore, the presence of TRIM21 and TNKS1BP in the NANOS2 interactome indicates that they could edit CNOT or NANOS2 with post-translational modifications. This, in turn, could modulate the putative repressive activity of the CNOT-NANOS2 complex and eventually affect the expression of NANOS2 targets in SSCs.

In summary, besides TRIM21 and TNKS1BP, CNOT subunits are the main NANOS2 interactors in mouse SSC lines. Importantly, since the NANOS2-CNOT1 interaction is essential for NANOS2 functions during early male germ cell development (Suzuki et al., 2012a), CNOT could also be critical for NANOS2 in SSC lines.

4.5.2 NANOS2-mRNA targets have a short half-life in SSC lines

After proving by IP-MS that NANOS2 mainly engages the CNOT complex in SSC lines, we hypothesised that this interaction could result in the degradation of NANOS2 targets. To address this, we measured mRNA stability in mouse SSCs by employing SLAM-seq. This is a novel metabolic labelling method, that quantifies mRNA half-life transcriptome-wide in live cells, also from low input material (Herzog et al., 2017).

Thus, I set-up SLAM-seq in SSC lines, and chased mRNA decay for 24 hours. Interestingly, SLAM-seq analysis revealed that the general mRNA half-life in SSCs is similar to the one in ESCs (224 and 229 minutes, respectively), despite the great difference in their doubling time (~5.6 days for SSCs (Kubota et al., 2004), and ~10-14 hours for ESCs (Pauklin et al., 2011)). Hence, these data suggest that mRNA turnover in SSCs is actively used and it is employed as much as rapidly dividing cells.

Next, we investigated the relationship between NANOS2 CRAC occupancy and mRNA half-life in SSCs. Strikingly, we discovered that the half-life of NANOS2 mRNA targets was significantly shorter (~180 minutes) compared to the average mRNA half-life in SSC lines (~224 minutes). Importantly, NANOS2 binding intensity correlated with higher mRNA instability.

Overall, these data strongly suggested that NANOS2 binding promotes the rapid turnover of its *bona fide* mRNA targets, in mouse SSC lines (Figure 4.12).

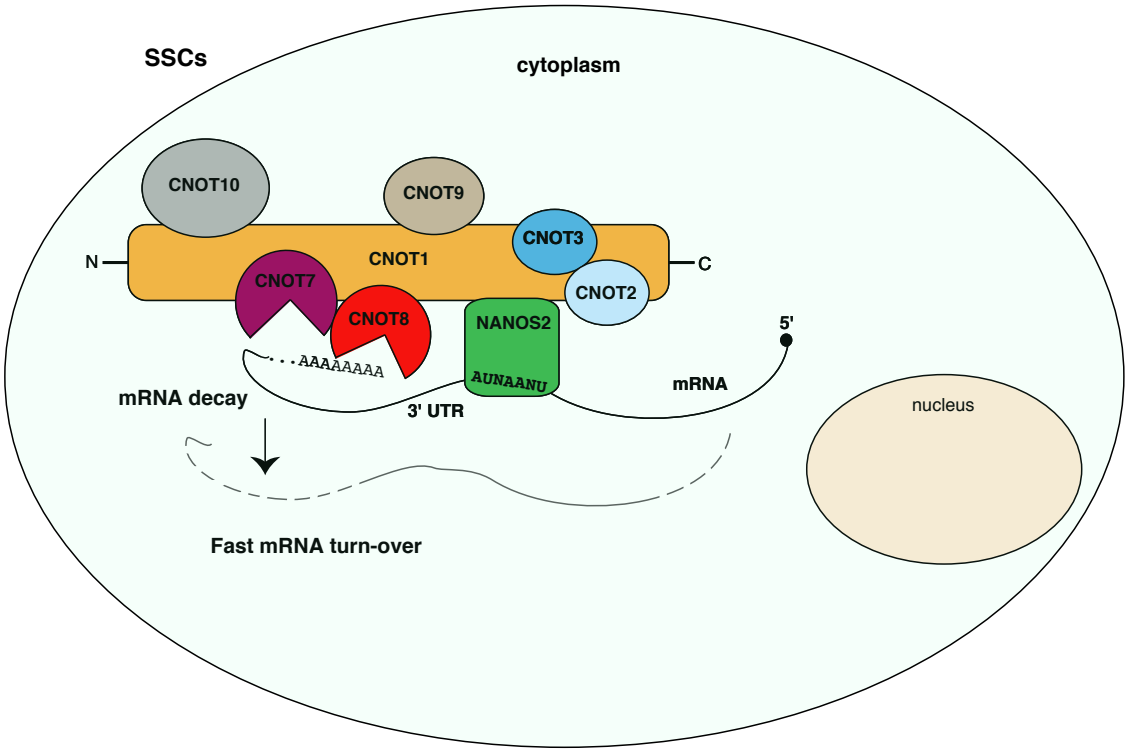


Figure 4.12. Model for NANOS2 molecular function in SSC lines.
The CNOT complex is the major NANOS2 interactor in SSC lines, as we determined by IP-MS. Since NANOS2 binds directly to the 3' UTR of mRNAs, it is likely that it recruits the CNOT complex to these transcripts and promote their decay. Indeed, NANOS2 binding triggers fast mRNA turn-over, as we assessed by performing SLAM-seq in SSC lines. We believe that the mechanism underlying mRNA decay is most likely deadenylation, with subsequent 5' decapping and 5'-3' degradation. However, we cannot exclude that also other types of deadenylation-independent repressions, mediated by CNOT, might occur.

It is worth noting that although we could calculate mRNA half-life for the majority of NANOS2 mRNA targets (~74%), our analysis does not take into account 26% of the NANOS2 CRAC targets. This is probably due to intrinsic limitations of SLAM-seq, such as the RNA labelling efficiency. Indeed, also in the original SLAM-seq protocol the authors reported the half-life of only ~8400 mRNAs in mESCs (Herzog et al., 2017). Thus, our speculations about the role

of NANOS2 in RNA decay are partially limited by the sensitivity of the SLAM-seq technique.

In general, attempts to investigate whether NANOS2 could induce RNA decay were made in different systems. For instance, NANOS2 was shown to deadenylate artificial poly(A)-RNAs *in vitro*, in the presence of CNOT, from Nanos2-overexpressing testes extracts (Suzuki et al., 2010). Similarly, also human NANOS proteins were shown to degrade mRNA reporters *in vitro* (Bhandari et al., 2014).
In SSC lines, NANOS2 was suggested to repress its targets through storage of 'dormant' mRNAs, promoted by condensation of mRNPs, or by their degradation (Zhou et al., 2015). However, our data showed that NANOS2 targets undergo fast RNA turnover, and thus, they are not dormant. Moreover, Zhou and colleagues assessed mRNAs repression by evaluating the upregulation of few putative NANOS2 targets (determined by native RIP) by RT-qPCR, in *Nanos2*$^{-/-}$ SSC lines (Zhou et al., 2015). However, when I analysed the half-life of the only two NANOS2 targets confirmed by CRAC (*Taf7l* and *Dmrt1*), I observed that they were significantly more stable than the average stability in SSC lines. Thus, although *Taf7l* and *Dmrt1* are important for SSCs differentiation and they are upregulated upon *Nanos2* loss, my data suggest that, in steady-state conditions, they do not undergo rapid turnover as a consequence of NANOS2 binding.

In the previous chapter, I described that NANOS2 bona fide targets identified by CRAC are mainly involved in cellular metabolism and include mTORC positive regulators. Strikingly, we found that the half-life of these mTORC regulators was significantly shorter than the average stability in SSC lines. Importantly, these data point to a role of NANOS2 in the negative regulation of the mTORC pathway. This could ensure mTORC repression in the self-renewing SSCs, thus protecting them from aberrant proliferation and stem cell depletion. At the same time, as soon as Nanos2 expression decreases in the

Miwi2^{+} and c-kit^{+} spermatogonia, NANOS2 targets would be released from its repression. Thus, following Nanos2 downregulation in differentiating spermatogonia, hundreds of metabolic mRNAs could be rapidly stabilised and possibly translated, in order to accelerate proliferation and promote a fast switch towards differentiation. Importantly, this post-transcriptional mechanism could provide a flexible means to ensure from one side maintenance of the self-renewal ability, and, from the other side, rapid differentiation. Although we already observed an upregulation of NANOS2 targets in differentiating spermatogonia (see Chapter 3), it would be interesting to assess whether they are also upregulated in *Nanos2*^{cKO} SSC lines.

Thus, NANOS2 binding induces rapid RNA turnover by shortening mRNA half-life in SSC lines. We showed that in the same cells, NANOS2 mainly interacts with the CNOT complex, which is involved in mRNAs deadenylation, the most rate-limiting step of RNA degradation. Thus, it is likely that NANOS2 mRNAs targets undergo CNOT-mediated deadenylation, which could be followed by 5' decapping and subsequent 5'-3' degradation. However, it is known that CNOT can destabilise mRNAs also via deadenylation-independent mechanisms (Collart and Panasenko, 2012; Cooke et al., 2010). Moreover, NANOS proteins were also shown to mediate translational repression of mRNAs which lack poly(A) tails, both in *Drosophila* and *Xenopus* (Chagnovich and Lehmann, 2001; Lai et al., 2011).

In mouse, attempts to investigate whether NANOS2 was responsible for the deadenylation of its targets, in complex with CNOT, were made by immunoprecipitating NANOS2 from Nanos2-overexpressing testes, and assessing the poly(A) length of artificial mRNAs *in vitro* (Suzuki et al., 2010). The investigation of mRNAs poly(A) length *in vivo* was impaired due to the low sensitivity of the technique used, in *Nanos2*$^{-/-}$ male gonocytes (Suzuki et al., 2010). Thus, it is currently unclear whether NANOS2, in complex with CNOT, would promote mRNA decay through deadenylation. To address this long-standing issue, we could measure the precise length of mRNAs poly(A) tails

in SSC lines and assess whether NANOS2 targets have indeed shorter tails. To do so, we could employ a high-throughput method called TAIL-seq (Chang et al., 2014), which now can also be applied to a low amount of material, such as a few hundred of mouse oocytes (Morgan et al., 2017).

It is also known that the CNOT complex can repress mRNAs through miRNAs (Inada and Makino, 2014; Shirai et al., 2014) and we determined that NANOS2 can also bind to miRNAs, although to a minor extent (see Chapter 3). Since one miRNA is thought to regulate hundreds of mRNAs (Bartel, 2009; Friedman et al., 2009), it is possible that NANOS2-miRNAs-CNOT complexes could trigger the decay of hundreds of transcripts in SSCs.

Moreover, the CNOT complex is involved in translational repression, also independently of deadenylation (Collart and Panasenko, 2012; Cooke et al., 2010). The precise mechanisms underlying its translational repression activity are poorly understood. However, this might occur through the interaction between CNOT and Dhh1, which is part of the 5' decapping complex (Collart and Panasenko, 2012; Maillet and Collart, 2002).
Additionally, NANOS is considered a translational repressor in different organisms (Bhandari et al., 2014; De Keuckelaere et al., 2018). Notably, human NANOS proteins can repress the translation of a reporter gene without affecting its mRNA levels *in vitro* (Bhandari et al., 2014). Thus, it is possible that NANOS2 targets in SSCs could also undergo translational repression, besides mRNA instability.
In mouse SSCs, sucrose gradient experiments determined that NANOS2 is mainly associated with the RNPs (ribonucleoproteins) and the light monosome fractions, suggesting that NANOS2 mRNAs targets should not be actively translated (Barrios et al., 2010; Zhou et al., 2015). Furthermore, upon Nanos2 deletion in SSC lines, there is a general increase of polysome-associated RNAs, suggesting that NANOS2 is a translational repressor (Zhou et al., 2015). However, this might also be an indirect effect of Nanos2 loss, and we

do not know whether NANOS2 really represses its bona fide targets at the translational level. Thus, to clearly define whether NANOS2 exclusively affects the RNA stability of its targets or also their translation, we could perform ribosome profiling (Reid et al., 2015) from *Nanos2*^{TAG} SSC lines. By doing so, we could compare the translation efficiency (TE) of NANOS2 targets with the general TE in SSCs.

In summary, it would be interesting to clarify whether NANOS2 targets are indeed degraded as a result of mRNA deadenylation, and also whether they undergo other types of post-transcriptional regulation, such as translational repression, besides mRNA decay.

4.5.3 mRNAs with a short half-life are enriched in NANOS2 motif

mRNA degradation is controlled by different factors, among which *cis*-elements in their 3' UTR, such as the 'ARE' (see Chapter 1). Therefore, we assessed whether the instability of NANOS2 mRNA targets was also related to the presence of particular sequences in their 3' UTR. Strikingly, we discovered that mRNAs with short half-life contained few motifs compatible with the 'AUNAANU' NANOS2 motif, previously identified by CRAC in SSC lines. Thus, these data provided the NANOS2 motif with a functional effect, since they suggest that it induces mRNA instability.

However, the motifs identified by Sylamer analysis were only moderately enriched in the 3' UTRs analysed. This could be explained by the relatively small size of the 3' UTRs analysed (~1000), which might reduce the statistical power of the enrichment analysis performed.
Moreover, the motifs identified by Sylamer represent variants (e.g., 'AUGAAUU') of the original NANOS2 motif ('AUNAANU'). Since the 'AUNAANU' did not correlate with NANOS2 binding intensity (see Chapter 3),

it is possible that, instead, these specific variants correlate better with high CRAC signal.

Furthermore, NANOS2 affinity towards RNA might also change depending on the slight nucleotide differences present in these motifs. It would be interesting to measure this by employing EMSA *in vitro* and using these specific motifs as substrates. Moreover, to further demonstrate that the presence of these motifs induces higher mRNA instability, we could reconstitute NANOS2-CNOT complexes *in vitro*, and test whether the addition RNA substrates with these motifs affects their mRNA degradation rate, by using a deadenylation assay (Morita et al., 2007).

Eventually, it would also be interesting to investigate whether NANOS2 targets with these motifs are enriched in specific biological processes, within the broad category of cellular metabolism.

In summary, we demonstrated that NANOS2 RNA-binding is associated with mRNA instability and that this is true for the 74% of its bona fide targets, identified by CRAC. This instability most likely occurs as a result of the NANOS2-CNOT interaction, since CNOT is the major NANOS2 interactor in SSC lines and the most critical deadenylase in eukaryotes.

Moreover, our data suggest that NANOS2 can select its targets and promote their instability, through the recognition of a novel *cis*-element in the 3' UTR, 'AUNAANU'.

Chapter 5. Conclusions

Undifferentiated spermatogonia have been studied for decades according to their morphology and topology in mouse testes (Phillips et al., 2010; Rooij, 2017). Then, the development of spermatogonial transplantation assays allowed the functional evaluation of their self-renewal capacity (Brinster and Zimmermann, 1994; Helsel and Oatley, 2017). Later on, several genes were identified as being specifically expressed in sub-groups of undifferentiated spermatogonia, where SSCs were considered to be enriched (Buaas et al., 2004; Costoya et al., 2004; Hofmann et al., 2005; Rooij, 2017; Sun et al., 2015; Suzuki et al., 2009; Yoshida et al., 2004). Some of these genes have been used as markers to follow the fate of SSCs, with lineage tracing experiments and live imaging, both under the steady-state and during regeneration (Nakagawa et al., 2007, 2010). These and other studies led scientists to conceive two different SSCs self-renewal models (Hara et al., 2014; Helsel et al., 2017; Nakagawa et al., 2010), which both challenged the classic 'A_s model' (Huckins, 1971; Oakberg, 1971) (see Chapter 1). Therefore, during the past ten years, much progress has been made in the field, which increased our understanding of SSCs' behaviour in mouse. However, the molecular mechanisms underlying SSCs self-renewal are still poorly understood, since the scarcity of SSCs *in vivo* has hampered their molecular characterisation for decades. Importantly, the development of an *in vitro* SSCs culture system in 2003 (Kanatsu-Shinohara et al., 2003) allowed expanding a large amount of cells, and thus provided an invaluable tool for molecular and biochemical studies in SSC lines (Takashima and Shinohara, 2018). By taking advantage of this SSCs culture system, in this work, I have investigated the molecular function of the RBP NANOS2, which is essential for mouse fertility and SSCs self-renewal (Sada et al., 2009; Tsuda et al., 2003). In particular, I employed a *Nanos2*^{TAG} mouse model to derive SSC lines and expand them *in vitro* in order to perform biochemical and molecular analyses.

To determine the NANOS2-RNA interactome accurately, we decided to use CRAC, which is one of the most stringent CLIP methods (Chapter 1, section 1.9). However, CRAC was originally designed for yeast cultures (Granneman et al., 2009a), which can provide grams of cells within hours. On the contrary, SSC lines proliferate slowly in culture and also express low levels of NANOS2, indicating that performing CRAC in these cells would have been a demanding task. However, I have successfully applied this technique to SSC lines, and this work is, to my knowledge, only the second one which employed CRAC in cells lines derived from adult mammalian tissues (Tebaldi et al., 2018).

Importantly, CRAC analysis revealed unprecedented findings concerning NANOS2 molecular interaction with RNA. First of all, we determined that NANOS2 binds directly to RNA in mouse SSC lines, which had not been unambiguously demonstrated before (see Chapter 3). In *Drosophila* NANOS-RNA binding is mediated by the RBP PUMILIO (De Keuckelaere et al., 2018), and in mouse male gonocytes NANOS2 is thought to act together with the RBP DND1 to repress specific RNAs (Suzuki et al., 2016). Thus, my data suggested that NANOS2 adopts a distinct mode of RNA-binding in mouse SSCs, which is independent of other RBPs. Additionally, this knowledge could be useful to better investigate the RNA-binding ability of other NANOS paralogs and orthologs.

Moreover, CRAC analysis showed that NANOS2 binds to RNA with sequence specificity, which also constitutes a new finding. Furthermore, we determined that NANOS2 binds to the 3' UTR of mRNAs in SSC lines and that these transcripts are involved in cellular metabolism. So far, only a handful of putative mRNAs were described to associate with NANOS2 in SSC lines (Zhou et al., 2015), and these were all related to spermatogenesis, thus suggesting that NANOS2 maintained SSCs self-renewal through the repression of spermatogenesis-specific transcripts. However, my data indicate that NANOS2 would rather repress metabolic mRNAs in SSC lines.

Among NANOS2 CRAC targets, I also identified a few positive mTORC regulators, which suggested that NANOS2 could repress the mTOR pathway

in SSC lines. Although the interplay between NANOS2 and mTOR was already shown in these cell lines (Zhou et al., 2015), this was demonstrated exclusively at the protein level, by virtue of their physical interaction.

To clarify how NANOS2 could regulate the fate of these metabolic mRNAs, we sought to determine its protein interactome by performing IP-MS. Importantly, we discovered that the CNOT proteins were the most robust and confident NANOS2 interactors in mouse SSC lines. Thus, my data reinforced the notion that the CNOT complex is a conserved NANOS2 interactor (Bhandari et al., 2014; Kadyrova et al., 2007; Suzuki et al., 2010), and also suggested that NANOS2-CNOT interaction could be essential for NANSO2 function in post-natal SSCs, as it was shown for male gonocytes (Suzuki et al., 2012a).

Hence, we hypothesised that NANOS2 could recruit CNOT to its mRNAs targets and promote their degradation. To address that, we employed SLAM-seq (Herzog et al., 2017) and measured mRNA stability transcriptome-wide in SSC lines. Importantly, we discovered that NANOS2 targets are unstable, compared to the average stability in SSCs. Therefore, although it was already hypothesised that NANOS2 could repress transcripts through RNA decay in SSCs (Zhou et al., 2015), we addressed this speculation by providing accurate experimental evidence, in SSC lines.

Overall, my data demonstrated a link between NANOS2 and the RNA decay pathway, in SSC lines, at least in two independent ways: first of all, by identifying CNOT by IP-MS and then, by proving a correlation between NANOS2 RNA-binding and higher mRNA instability, by SLAM-seq.

Thus, these findings improved our comprehension of NANOS2 molecular function in SSC lines and supported a model where NANOS2 could act as a repressor of hundreds of mRNAs, by promoting their instability, at the post-transcriptional level, in SSCs (Figure 5.1). Moreover, since unstable transcripts are generally associated with regulatory functions (Herzog et al., 2017; Schwanhäusser et al., 2011), NANOS2 targets could contribute to the regulation of critical processes, such as the transition from self-renewing SSCs to differentiating spermatogonia. If NANOS2 really repressed its targets in

SSCs, we would expect that this repression was released upon NANOS2 downregulation in differentiating spermatogonia. To address this, we analysed the transcriptomic changes across adult spermatogonia development. Interestingly, we discovered that one-third of NANOS2 *bona fide* targets were moderately upregulated during the transition from self-renewing SSCs (GFRα1$^+$, NANOS2$^+$) to transit-amplifying spermatogonia (MIWI2$^+$) (Figure 5.1).

Post-transcriptional regulation of gene expression is extensively employed by cells during critical development processes, as it provides a dynamic means to modify cellular programs (see Chapter 1). Our data support a model where the association between NANOS2 and CNOT could induce the degradation of hundreds of transcripts at the post-transcriptional level in SSCs. However, following the downregulation of NANOS2 in differentiating spermatogonia, this post-transcriptional mechanism could be switched off promptly, and NANOS2 targets could be stabilised. Thus, NANOS2 could act in the GFRα1$^+$ SSCs pool by suppressing a critical developmental decision, which is the first step towards differentiation. Since this process would occur by employing a flexible regulative mechanism, in the absence of NANOS2, there could be a rapid resumption of NANOS2 targets expression. This, in turn, would promote a faster metabolism and thus support the higher proliferation rates required by the MIWI2$^+$ spermatogonia (Figure 5.1).

Therefore, by doing so, NANOS2 could actively participate in the maintenance of cellular quiescence and self-renewal of GFrα1$^+$ SSCs, and allow, at the same time, their rapid switch towards differentiation, as soon as NANOS2 levels decrease.

In conclusion, the results presented herein revealed unprecedented insight into NANOS2 mechanistic function in mouse SSCs.

Moreover, they improved our knowledge about the possible molecular basis of SSCs maintenance, a crucial process which supplies an endless source of cells to spermatogenesis, throughout adult life.

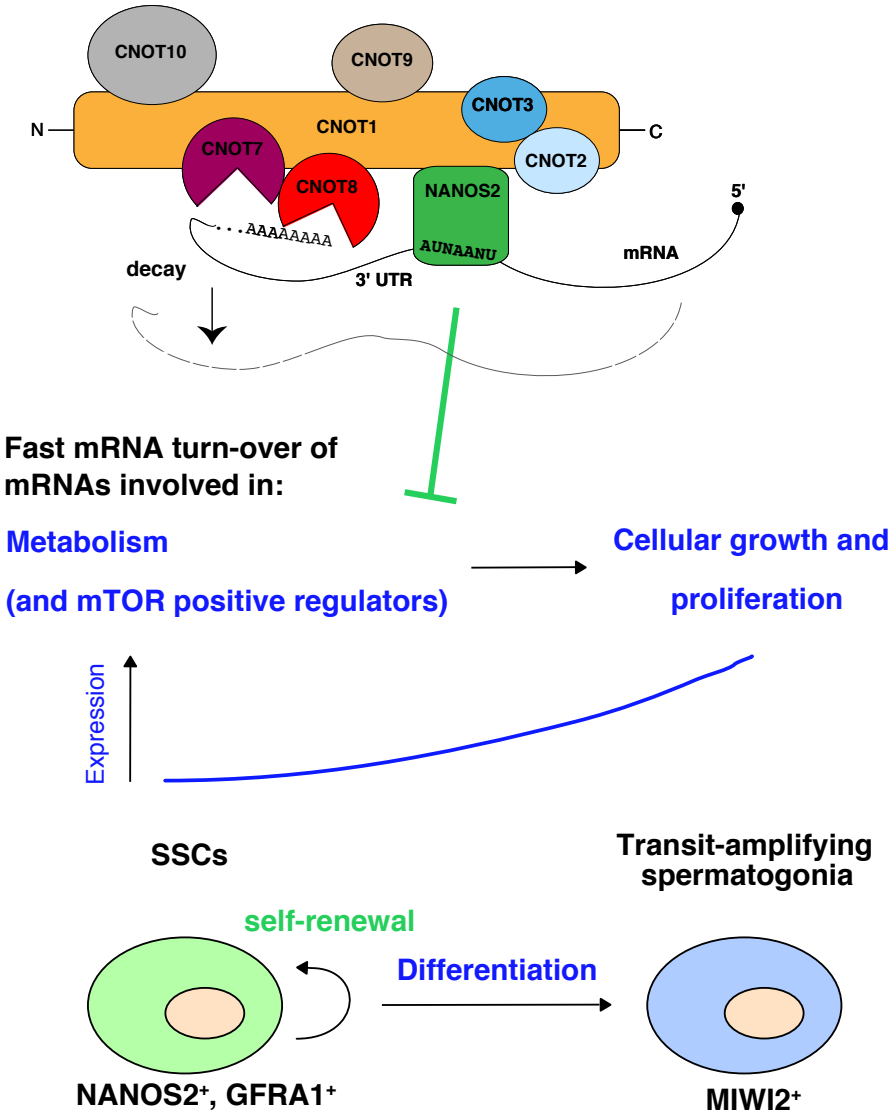


Figure 5.1. Possible working model for NANOS2 in SSCs.

We showed that NANOS2 binds to the 3' UTR of mRNAs with sequence specificity in SSC lines. There, it also associates with the CCR4-NOT complex and promotes mRNA instability. The mRNAs bound by NANOS2 are enriched in cellular metabolic processes, including few positive mTORC regulators among the top targets. Self-renewing SSCs control metabolism strictly in order to maintain the balance between cellular quiescence and proliferation. Since one-third of NANOS2 targets are upregulated during the transition from Gfrα1$^+$ to Miwi2$^+$ spermatogonia, we propose a model where NANOS2 represses these transcripts in the Gfrα1$^+$ SSCs to preserve quiescence and the long-term self-renewing pool. Once NANOS2 is downregulated in the transit-amplifying Miwi2$^+$ spermatogonia, the stability of these metabolic

transcripts would increase, thus allowing the resumption of their expression. This would result in the positive regulation of cellular growth, which, in turn, would support the higher proliferation rates required by the transit-amplifying populations of spermatogonia.

References

Aponte, P.M., Bragt, M.P. a. V., Rooij, D.G.D., and Pelt, A.M.M.V. (2005). Spermatogonial stem cells: characteristics and experimental possibilities. APMIS *113*, 727–742.

Arrizabalaga, G., and Lehmann, R. (1999). A selective screen reveals discrete functional domains in Drosophila Nanos. Genetics *153*, 1825–1838.

Asaoka-Taguchi, M., Yamada, M., Nakamura, A., Hanyu, K., and Kobayashi, S. (1999). Maternal Pumilio acts together with Nanos in germline development in Drosophila embryos. Nat Cell Biol *1*, 431–437.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME Suite: tools for motif discovery and searching. Nucleic Acids Res *37*, W202–W208.

Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. Molecular Cell *46*, 674–690.

Barrios, F., Filipponi, D., Pellegrini, M., Paronetto, M.P., Siena, S.D., Geremia, R., Rossi, P., Felici, M.D., Jannini, E.A., and Dolci, S. (2010). Opposing effects of retinoic acid and FGF9 on Nanos2 expression and meiotic entry of mouse germ cells. J Cell Sci *123*, 871–880.

Bartel, D.P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. Cell *136*, 215–233.

Bawankar, P., Loh, B., Wohlbold, L., Schmidt, S., and Izaurralde, E. (2013). NOT10 and C2orf29/NOT11 form a conserved module of the CCR4-NOT complex that docks onto the NOT1 N-terminal domain. RNA Biol *10*, 228–244.

Becalska, A.N., Kim, Y.R., Belletier, N.G., Lerit, D.A., Sinsimer, K.S., and Gavis, E.R. (2011). Aubergine is a component of a nanos mRNA localization complex. Developmental Biology *349*, 46–52.

Beckmann, B.M., and Granneman, S. (2019). Probing the RNA-Binding Proteome from Yeast to Man: Major Advances and Challenges. In Yeast Systems Biology, S.G. Oliver, and J.I. Castrillo, eds. (New York, NY: Springer New York), pp. 213–231.

Beckmann, B.M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A.-M., Schwarzl, T., Curk, T., Foehr, S., Huber, W., et al. (2015). The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. Nature Communications *6*, 10127.

Bellve, A.R., Cavicchia, J.C., Millette, C.F., O'Brien, D.A., Bhatnagar, Y.M., and Dym, M. (1977). Spermatogenic cells of the prepuberal mouse: isolation and morphological characterization. The Journal of Cell Biology *74*, 68–85.

Bergsten, S.E., and Gavis, E.R. (1999). Role for mRNA localization in translational activation but not spatial restriction of nanos RNA. Development *126*, 659–669.

Berthet, C., Morera, A.-M., Asensio, M.-J., Chauvin, M.-A., Morel, A.-P., Dijoud, F., Magaud, J.-P., Durand, P., and Rouault, J.-P. (2004). CCR4-Associated Factor CAF1 Is an Essential Factor for Spermatogenesis. Mol Cell Biol *24*, 5808–5820.

Bevilacqua, P.C., Ritchey, L.E., Su, Z., and Assmann, S.M. (2016). Genome-Wide Analysis of RNA Secondary Structure. Annu. Rev. Genet. *50*, 235–266.

Bhandari, D., Raisch, T., Weichenrieder, O., Jonas, S., and Izaurralde, E. (2014). Structural basis for the Nanos-mediated recruitment of the CCR4–NOT complex and translational repression. Genes Dev. *28*, 888–901.

Blow, J.J., and Dutta, A. (2005). Preventing re-replication of chromosomal DNA. Nat Rev Mol Cell Biol *6*, 476–486.

Bresson, S., and Tollervey, D. (2018). Tailing Off: PABP and CNOT Generate Cycles of mRNA Deadenylation. Molecular Cell *70*, 987–988.

Brinster, R.L., and Zimmermann, J.W. (1994). Spermatogenesis following male germ-cell transplantation. Proceedings of the National Academy of Sciences *91*, 11298–11302.

Buaas, F.W., Kirsh, A.L., Sharma, M., McLean, D.J., Morris, J.L., Griswold, M.D., Rooij, D.G. de, and Braun, R.E. (2004). Plzf is required in adult male germ cells for stem cell self-renewal. Nature Genetics *36*, 647.

Bühler, M., Spies, N., Bartel, D.P., and Moazed, D. (2008). TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the Schizosaccharomyces pombe siRNA pathway. Nat Struct Mol Biol *15*, 1015–1023.

Cano, F., Miranda-Saavedra, D., and Lehner, P.J. (2010). RNA-binding E3 ubiquitin ligases: novel players in nucleic acid regulation. Biochemical Society Transactions *38*, 1621–1626.

Cano, F., Rapiteanu, R., Sebastiaan Winkler, G., and Lehner, P.J. (2015a). A non-proteolytic role for ubiquitin in deadenylation of MHC-I mRNA by the RNA-binding E3-ligase MEX-3C. Nat Commun *6*, 8670.

Cao, Z., Mao, X., and Luo, L. (2019). Germline Stem Cells Drive Ovary Regeneration in Zebrafish. Cell Reports *26*, 1709-1717.e3.

Carballo, E., Lai, W.S., and Blackshear, P.J. (1998a). Feedback inhibition of macrophage tumor necrosis factor-alpha production by tristetraprolin. Science *281*, 1001–1005.

Carballo, E., Lai, W.S., and Blackshear, P.J. (1998b). Feedback Inhibition of Macrophage Tumor Necrosis Factor-α Production by Tristetraprolin. Science *281*, 1001–1005.

Carrieri, C., Comazzetto, S., Grover, A., Morgan, M., Buness, A., Nerlov, C., and O'Carroll, D. (2017). A transit-amplifying population underpins the efficient regenerative capacity of the testis. The Journal of Experimental Medicine *214*, 1631–1641.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. Cell *149*, 1393–1406.

Castilho, R.M., Squarize, C.H., Chodosh, L.A., Williams, B.O., and Gutkind, J.S. (2009). mTOR Mediates Wnt-Induced Epidermal Stem Cell Exhaustion and Aging. Cell Stem Cell *5*, 279–289.

Chagnovich, D., and Lehmann, R. (2001). Poly(A)-independent regulation of maternal hunchback translation in the Drosophila embryo. Proceedings of the National Academy of Sciences *98*, 11359–11364.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. Cell *113*, 643–655.

Chan, F., Oatley, M.J., Kaucher, A.V., Yang, Q.-E., Bieberich, C.J., Shashikant, C.S., and Oatley, J.M. (2014). Functional and molecular features of the Id4+ germline stem cell population in mouse testes. Genes Dev. *28*, 1351–1362.

Chang, H., Lim, J., Ha, M., and Kim, V.N. (2014). TAIL-seq: Genome-wide Determination of Poly(A) Tail Length and 3′ End Modifications. Molecular Cell *53*, 1044–1052.

Chen, J., Rappsilber, J., Chiang, Y.-C., Russell, P., Mann, M., and Denis, C.L. (2001). Purification and characterization of the 1.0 MDa CCR4-NOT complex identifies two novel components of the complex11Edited by D. Draper. Journal of Molecular Biology *314*, 683–694.

Cheng, Y., Buffone, M.G., Kouadio, M., Goodheart, M., Page, D.C., Gerton, G.L., Davidson, I., and Wang, P.J. (2007). Abnormal Sperm in Mice Lacking the Taf7l Gene. Molecular and Cellular Biology *27*, 2582–2589.

Chiarini-Garcia, H., and Russell, L.D. (2001). High-Resolution Light Microscopic Characterization of Mouse Spermatogonia. Biol Reprod *65*, 1170–1178.

Cho, J., Chang, H., Kwon, S.C., Kim, B., Kim, Y., Choe, J., Ha, M., Kim, Y.K., and Kim, V.N. (2012). LIN28A Is a Suppressor of ER-Associated Translation in Embryonic Stem Cells. Cell *151*, 765–777.

Cho, P.F., Gamberi, C., Cho-Park, Y.A., Cho-Park, I.B., Lasko, P., and Sonenberg, N. (2006). Cap-Dependent Translational Inhibition Establishes Two Opposing Morphogen Gradients in Drosophila Embryos. Current Biology *16*, 2035–2041.

Choo, Y., and Klug, A. (1994). Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. Proceedings of the National Academy of Sciences *91*, 11168–11172.

Collart, M.A., and Panasenko, O.O. (2012). The Ccr4–Not complex. Gene *492*, 42–53.

Cooke, A., Prigge, A., and Wickens, M. (2010). Translational Repression by Deadenylases. J. Biol. Chem. *285*, 28506–28513.

Corbett, A.H. (2018). Post-transcriptional regulation of gene expression and human disease. Current Opinion in Cell Biology *52*, 96–104.

Costoya, J.A., Hobbs, R.M., Barna, M., Cattoretti, G., Manova, K., Sukhwani, M., Orwig, K.E., Wolgemuth, D.J., and Pandolfi, P.P. (2004). Essential role of Plzf in maintenance of spermatogonial stem cells. Nature Genetics *36*, 653.

Cox et al. (2014). Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. Mol Cell Proteomics.

Curtis, D., Treiber, D.K., Tao, F., Zamore, P.D., Williamson, J.R., and Lehmann, R. (1997). A CCHC metal-binding domain in Nanos is essential for translational regulation. EMBO J *16*, 834–843.

Dalby, B., and Glover, D. m. (1993). Discrete sequence elements control posterior pole accumulation and translational repression of maternal cyclin B RNA in Drosophila. The EMBO Journal *12*, 1219–1227.

De Keuckelaere, E., Hulpiau, P., Saeys, Y., Berx, G., and van Roy, F. (2018). Nanos genes and their role in development and beyond. Cell. Mol. Life Sci. *75*, 1929–1946.

Ding, Y., Chan, C.Y., and Lawrence, C.E. (2005). RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. RNA *11*, 1157–1166.

Dobin, A., Carrie A. D., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., and Gingeras T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15– 21.

Dodt, M., Roehr, J., Ahmed, R., and Dieterich, C. (2012). FLEXBAR—Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. Biology *1*, 895–905.

van Dongen, S., Abreu-Goodger, C., and Enright, A.J. (2008). Detecting microRNA binding and siRNA off-target effects from expression data. Nat Methods *5*, 1023–1025.

Dreyfuss, G., Kim, V.N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. Nature Reviews Molecular Cell Biology *3*, 195.

Dunleavy, J.E.M., O'Bryan, M.K., Stanton, P.G., and O'Donnell, L. (2019). The cytoskeleton in spermatogenesis. Reproduction R53–R72.

Elliott, D. (2003). Pathways of post-transcriptional gene regulation in mammalian germ cell development. Cytogenet Genome Res *103*, 210–216.

Farley, F.W., Soriano, P., Steffen, L.S., and Dymecki, S.M. (2000). Widespread recombinase expression using FLPeR (Flipper) mice. Genesis *28*, 106–110.

Forbes, A., and Lehmann, R. (1998). Nanos and Pumilio have critical roles in the development and function of Drosophila germline stem cells. Development *125*, 679-690.

França, L.R., Hess, R.A., Dufour, J.M., Hofmann, M.C., and Griswold, M.D. (2016). The Sertoli cell: one hundred fifty years of beauty and plasticity. Andrology *4*, 189–212.

Freeman, T.C., Goldovsky, L., Brosch, M., Dongen, S. van, Mazière, P., Grocock, R.J., Freilich, S., Thornton, J., and Enright, A.J. (2007). Construction, Visualisation, and Clustering of Transcription Networks from Microarray Expression Data. PLOS Computational Biology *3*, e206.

Friedman, R.C., Farh, K.K.-H., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res *19*, 92–105.

Gan, B., and DePinho, R. (2009). mTORC1 signaling governs hematopoietic stem cell quiescence. Cell Cycle *8*, 1003–1006.

Garces, R.G., Gillon, W., and Pai, E.F. (2007). Atomic model of human Rcd-1 reveals an armadillo-like-repeat protein with in vitro nucleic acid binding properties. Protein Sci *16*, 176–188.

Gavis, E.R., and Lehmann, R. (1992). Localization of nanos RNA controls embryonic polarity. Cell *71*, 301–313.

Gavis, E.R., and Lehmann, R. (1994). Translational regulation of nanos by RNA localization. Nature *369*, 315–318.

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. Nature Reviews Genetics *15*, 829–845.

Gibson, B.A., and Kraus, W.L. (2012). New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. Nature Reviews Molecular Cell Biology *13*, 411–424.

Gilboa, L., and Lehmann, R. (2004). Repression of Primordial Germ Cell Differentiation Parallels Germ Line Stem Cell Maintenance. Current Biology *14*, 981–986.

Giudice, G., Sánchez-Cabo, F., Torroja, C., and Lara-Pezzi, E. (2016). ATtRACT—a database of RNA-binding proteins and associated motifs. Database (Oxford) *2016*.

Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. PNAS *106*, 9613–9618.

Greenberg, J.R. (1979). Ultraviolet light-induced crosslinking of mRNA to proteins. Nucleic Acids Res *6*, 715–732.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L. (2008). The Vienna RNA Websuite. Nucleic Acids Res *36*, W70–W74.

Guan, K., Nayernia, K., Maier, L.S., Wagner, S., Dressel, R., Lee, J.H., Nolte, J., Wolf, F., Li, M., Engel, W., et al. (2006). Pluripotency of spermatogonial stem cells from adult mouse testis. Nature *440*, 1199.

Guo, J., Grow, E.J., Yi, C., Mlcochova, H., Maher, G.J., Lindskog, C., Murphy, P.J., Wike, C.L., Carrell, D.T., Goriely, A., et al. (2017). Chromatin and Single-Cell RNA-Seq Profiling Reveal Dynamic Signaling and Metabolic Transitions during Human Spermatogonial Stem Cell Development. Cell Stem Cell *21*, 533-546.e6.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., et al. (2010). PAR-CliP - A Method to Identify Transcriptome-wide the Binding Sites of RNA Binding Proteins. J Vis Exp.

Haglund, K., Nezis, I.P., and Stenmark, H. (2011). Structure and functions of stable intercellular bridges formed by incomplete cytokinesis during development. Commun Integr Biol *4*, 1–9.

Haimovich, G., Choder, M., Singer, R.H., and Trcek, T. (2013). The fate of the messenger is pre-determined: A new model for regulation of gene expression. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1829*, 643–653.

Hamburgh, M.E., Curr, K.A., Monaghan, M., Rao, V.R., Tripathi, S., Preston, B.D., Sarafianos, S., Arnold, E., Darden, T., and Prasad, V.R. (2006). Structural Determinants of Slippage-mediated Mutations by Human Immunodeficiency Virus Type 1 Reverse Transcriptase. J. Biol. Chem. *281*, 7421–7428.

Hara, K., Nakagawa, T., Enomoto, H., Suzuki, M., Yamamoto, M., Simons, B.D., and Yoshida, S. (2014). Mouse Spermatogenic Stem Cells Continually Interconvert between Equipotent Singly Isolated and Syncytial States. Cell Stem Cell *14*, 658–672.

Haraguchi, S., Tsuda, M., Kitajima, S., Sasaoka, Y., Nomura-Kitabayashi, A., Kurokawa, K., and Saga, Y. (2003). nanos1: a mouse nanos gene expressed in the central nervous system is dispensable for normal development. Mechanisms of Development *120*, 721–731.

Hashimoto, H., Hara, K., Hishiki, A., Kawaguchi, S., Shichijo, N., Nakamura, K., Unzai, S., Tamaru, Y., Shimizu, T., and Sato, M. (2010a). Crystal structure of zinc-finger domain of Nanos and its functional implications. EMBO Rep *11*, 848–853.

Hashimoto, H., Hara, K., Hishiki, A., Kawaguchi, S., Shichijo, N., Nakamura, K., Unzai, S., Tamaru, Y., Shimizu, T., and Sato, M. (2010b). Crystal structure of zinc-finger domain of Nanos and its functional implications. EMBO Rep *11*, 848–853.

Hayashi, K., Lopes, S.M.C. de S., and Surani, M.A. (2007). Germ Cell Specification in Mice. Science *316*, 394–396.

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells. Cell *146*, 519–532.

Hayashi, Y., Hayashi, M., and Kobayashi, S. (2004). Nanos suppresses somatic cell fate in Drosophila germ line. PNAS *101*, 10338–10342.

He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D.C., Wilusz, J.E., Garcia, B.A., and Bonasio, R. (2016). High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. Molecular Cell *64*, 416–430.

Helsel, A.R., and Oatley, J.M. (2017). Transplantation as a Quantitative Assay to Study Mammalian Male Germline Stem Cells. In Germline Stem Cells, M. Buszczak, ed. (New York, NY: Springer New York), pp. 155–172.

Helsel, A.R., Yang, Q.-E., Oatley, M.J., Lord, T., Sablitzky, F., and Oatley, J.M. (2017). ID4 levels dictate the stem cell state in mouse spermatogonia. Development *144*, 624–634.

Helwak, A., and Tollervey, D. (2014). Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH). Nature Protocols *9*, 711–728.

Hentze, M.W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins. Nature Reviews Molecular Cell Biology *19*, 327–341.

Herzog, V.A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T.R., Wlotzka, W., von Haeseler, A., Zuber, J., and Ameres, S.L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. Nat Methods *14*, 1198–1204.

Hildebrandt, A., Alanis-Lobato, G., Voigt, A., Zarnack, K., Andrade-Navarro, M.A., Beli, P., and König, J. (2017). Interaction profiling of RNA-binding ubiquitin ligases reveals a link between posttranscriptional regulation and the ubiquitin system. Sci Rep *7*, 16582.

Hobbs, R.M., Seandel, M., Falciatori, I., Rafii, S., and Pandolfi, P.P. (2010). Plzf Regulates Germline Progenitor Self-Renewal by Opposing mTORC1. Cell *142*, 468–479.

Hockensmith, J.W., Kubasek, W.L., Vorachek, W.R., and Hippel, P.H. von (1986). Laser cross-linking of nucleic acids to proteins. Methodology and first applications to the phage T4 DNA replication system. J. Biol. Chem. *261*, 3512–3518.

Hofacker, I.L. (2003). Vienna RNA secondary structure server. Nucleic Acids Research *31*, 3429–3431.

Hofmann, M.-C., Braydich-Stolle, L., and Dym, M. (2005). Isolation of male germ-line stem cells; influence of GDNF. Developmental Biology *279*, 114–124.

Houseley, J., and Tollervey, D. (2009). The Many Pathways of RNA Degradation. Cell *136*, 763–776.

Hua, C., Zhao, G., Li, Y., and Bie, L. (2014). Minichromosome Maintenance (MCM) Family as potential diagnostic and prognostic tumor markers for human gliomas. BMC Cancer *14*, 526.

Hubner, N.C., Bird, A.W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., and Mann, M. (2010). Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. The Journal of Cell Biology *189*, 739–754.

Huckins, C. (1971). The spermatogonial stem cell population in adult rats. I. Their morphology, proliferation and maturation. The Anatomical Record *169*, 533–557.

Hülskamp, M., Schröder, C., Pfeifle, C., Jäckle, H., and Tautz, D. (1989). Posterior segmentation of the Drosophila embryo in the absence of a maternal posterior organizer gene. Nature *338*, 629–632.

Inada, T., and Makino, S. (2014). Novel roles of the multi-functional CCR4-NOT complex in post-transcriptional regulation. Front. Genet. *5*.

Irish, V., Lehmann, R., and Akam, M. (1989). The Drosophila posterior-group gene nanos functions by repressing hunchback activity. Nature *338*, 646–648.

Ishii, K., Kanatsu-Shinohara, M., Toyokuni, S., and Shinohara, T. (2012). FGF2 mediates mouse spermatogonial stem cell self-renewal via upregulation of Etv5 and Bcl6b through MAP2K1 activation. Development *139*, 1734–1743.

Ito, K., Takahashi, A., Morita, M., Suzuki, T., and Yamamoto, T. (2011). The role of the CNOT1 subunit of the CCR4-NOT complex in mRNA deadenylation and cell viability. Protein Cell *2*, 755–763.

Jacobs, E., Mills, J.D., and Janitz, M. (2012). The Role of RNA Structure in Posttranscriptional Regulation of Gene Expression. Journal of Genetics and Genomics *39*, 535–543.

Jain, R.A., and Gavis, E.R. (2008). The Drosophila hnRNP M homolog Rumpelstiltskin regulates nanos mRNA localization. Development *135*, 973–982.

Jiang, X.-H., Bukhari, I., Zheng, W., Yin, S., Wang, Z., Cooke, H.J., and Shi, Q.-H. (2014). Blood-testis barrier and spermatogenesis: lessons from genetically-modified mice. Asian J Androl *16*, 572–580.

Joly, W., Chartier, A., Rojas-Rios, P., Busseau, I., and Simonelig, M. (2013). The CCR4 Deadenylase Acts with Nanos and Pumilio in the Fine-Tuning of Mei-P26 Expression to Promote Germline Stem Cell Self-Renewal. Stem Cell Reports *1*, 411–424.

Kadyrova, L.Y., Habara, Y., Lee, T.H., and Wharton, R.P. (2007). Translational control of maternal Cyclin B mRNA by Nanos in the Drosophila germline. Development *134*, 1519–1527.

Kanatsu-Shinohara, M. (2005). Genetic and epigenetic properties of mouse male germline stem cells during long-term culture. Development *132*, 4155–4163.

Kanatsu-Shinohara, M., and Shinohara, T. (2013). Spermatogonial Stem Cell Self-Renewal and Development. Annual Review of Cell and Developmental Biology *29*, 163–187.

Kanatsu-Shinohara, M., Ogonuki, N., Inoue, K., Miki, H., Ogura, A., Toyokuni, S., and Shinohara, T. (2003). Long-Term Proliferation in Culture and Germline Transmission of Mouse Male Germline Stem Cells1. Biology of Reproduction *69*, 612–616.

Kanatsu-Shinohara, M., Inoue, K., Lee, J., Yoshimoto, M., Ogonuki, N., Miki, H., Baba, S., Kato, T., Kazuki, Y., Toyokuni, S., et al. (2004). Generation of Pluripotent Stem Cells from Neonatal Mouse Testis. Cell *119*, 1001–1012.

Kanatsu-Shinohara, M., Miki, H., Inoue, K., Ogonuki, N., Toyokuni, S., Ogura, A., and Shinohara, T. (2005). Long-Term Culture of Mouse Male Germline Stem Cells Under Serum-or Feeder-Free Conditions. Biol Reprod *72*, 985–991.

Kanatsu-Shinohara, M., Takehashi, M., and Shinohara, T. (2008). Brief History, Pitfalls, and Prospects of Mammalian Spermatogonial Stem Cell Research. Cold Spring Harb Symp Quant Biol *73*, 17–23.

Kanatsu-Shinohara, M., Ogonuki, N., Matoba, S., Morimoto, H., Ogura, A., and Shinohara, T. (2014). Improved Serum- and Feeder-Free Culture of Mouse Germline Stem Cells. Biol Reprod *91*.

Kanatsu-Shinohara, M., Tanaka, T., Ogonuki, N., Ogura, A., Morimoto, H., Cheng, P.F., Eisenman, R.N., Trumpp, A., and Shinohara, T. (2016). Myc/Mycn-mediated glycolysis enhances mouse spermatogonial stem cell self-renewal. Genes Dev. *30*, 2637–2648.

Kato, Y., Katsuki, T., Kokubo, H., Masuda, A., and Saga, Y. (2016). *Dazl* is a target RNA suppressed by mammalian NANOS2 in sexually differentiating male germ cells. Nature Communications *7*, ncomms11272.

Ke, Y., Zhang, J., Lv, X., Zeng, X., and Ba, X. (2019). Novel insights into PARPs in gene expression: regulation of RNA metabolism. Cell. Mol. Life Sci.

Keene, J.D., Komisarow, J.M., and Friedersdorf, M.B. (2006). RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. Nat Protoc *1*, 302–307.

Keskeny, C., Raisch, T., Sgromo, A., Igreja, C., Bhandari, D., Weichenrieder, O., and Izaurralde, E. (2019). A conserved CAF40-binding motif in metazoan NOT4 mediates association with the CCR4–NOT complex. Genes Dev. *33*, 236–252.

Kiebler, M.A., Scheiffele, P., and Ule, J. (2013). What, where, and when: the importance of post-transcriptional regulation in the brain. Front. Neurosci. *7*.

Kimmins, S., Kotaja, N., Davidson, I., and Sassone-Corsi, P. (2004). Testis-specific transcription mechanisms promoting male germ-cell differentiation. Reproduction *128*, 5–12.

Kobayashi, S., Yamada, M., Asaoka, M., and Kitamura, T. (1996). Essential role of the posterior morphogen nanos for germline development in Drosophila. Nature *380*, 708–711.

Komai, Y., Tanaka, T., Tokuyama, Y., Yanai, H., Ohe, S., Omachi, T., Atsumi, N., Yoshida, N., Kumano, K., Hisha, H., et al. (2015). Bmi1 expression in long-term germ stem cells. Sci Rep *4*, 6175.

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nature Structural & Molecular Biology *17*, 909–915.

Köprunner, M., Thisse, C., Thisse, B., and Raz, E. (2001). A zebrafish nanos-related gene is essential for the development of primordial germ cells. Genes Dev. *15*, 2877–2885.

Koubova, J., Menke, D.B., Zhou, Q., Capel, B., Griswold, M.D., and Page, D.C. (2006). Retinoic acid regulates sex-specific timing of meiotic initiation in mice. Proceedings of the National Academy of Sciences *103*, 2474–2479.

Kramer, K., Sachsenberg, T., Beckmann, B.M., Qamar, S., Boon, K.-L., Hentze, M.W., Kohlbacher, O., and Urlaub, H. (2014). Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. Nature Methods *11*, 1064–1070.

Kubota, H., Avarbock, M.R., and Brinster, R.L. (2004). Growth factors essential for self-renewal and expansion of mouse spermatogonial stem cells. Proceedings of the National Academy of Sciences *101*, 16489–16494.

Kuersten, S., and Goodwin, E.B. (2003). The power of the 3′ UTR: translational control and development. Nature Reviews Genetics *4*, 626–637.

Kusz-Zamelczyk, K., Sajek, M., Spik, A., Glazar, R., Jędrzejczak, P., Latos-Bieleńska, A., Kotecki, M., Pawelczyk, L., and Jaruzelska, J. (2013). Mutations of NANOS1, a human homologue of the Drosophila morphogen, are associated with a lack of germ cells in testes or severe oligo-astheno-teratozoospermia. Journal of Medical Genetics *50*, 187–193.

Kwok, C.K., Tang, Y., Assmann, S.M., and Bevilacqua, P.C. (2015). The RNA structurome: transcriptome-wide structure probing with next-generation sequencing. Trends in Biochemical Sciences *40*, 221–232.

Lai, F., Zhou, Y., Luo, X., Fox, J., and King, M.L. (2011). Nanos1 functions as a translational repressor in the Xenopus germline. Mechanisms of Development *128*, 153–163.

Lal, L., Li, Y., Smith, J., Sassano, A., Uddin, S., Parmar, S., Tallman, M.S., Minucci, S., Hay, N., and Platanias, L.C. (2005). Activation of the p70 S6 kinase by all-trans-retinoic acid in acute promyelocytic leukemia cells. Blood *105*, 1669–1677.

Laplante, M., and Sabatini, D.M. (2012). mTOR Signaling in Growth Control and Disease. Cell *149*, 274–293.

Lapointe, C.P., Wilinski, D., Saunders, H.A.J., and Wickens, M. (2015). Protein-RNA networks revealed through covalent RNA marks. Nature Methods *12*, 1163–1170.

Lau, N.-C., Kolkman, A., van Schaik, F.M.A., Mulder, K.W., Pijnappel, W.W.M.P., Heck, A.J.R., and Timmers, H.T.M. (2009). Human Ccr4–Not complexes contain variable deadenylase subunits. Biochemical Journal *422*, 443–453.

Lawson, K.A., Dunn, N.R., Roelen, B.A.J., Zeinstra, L.M., Davis, A.M., Wright, C.V.E., Korving, J.P.W.F.M., and Hogan, B.L.M. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. Genes Dev *13*, 424–436.

Lee, F.C.Y., and Ule, J. (2018). Advances in CLIP Technologies for Studies of Protein-RNA Interactions. Molecular Cell *69*, 354–369.

Lee, J., Kanatsu-Shinohara, M., Inoue, K., Ogonuki, N., Miki, H., Toyokuni, S., Kimura, T., Nakano, T., Ogura, A., and Shinohara, T. (2007). Akt mediates self-renewal division of mouse spermatogonial stem cells. Development *134*, 1853–1859.

Lee, M., Kim, B., and Kim, V.N. (2014). Emerging Roles of RNA Modification: m6A and U-Tail. Cell *158*, 980–987.

Lehmann, R., and Nusslein-Volhard, C. (1991). The maternal gene nanos has a central role in posterior pattern formation of the Drosophila embryo. Development *112*, 679–691.

Lerner, M.R., and Steitz, J.A. (1979). Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. PNAS *76*, 5495–5499.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., et al. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature *456*, 464–469.

Lim, J., Ha, M., Chang, H., Kwon, S.C., Simanshu, D.K., Patel, D.J., and Kim, V.N. (2014). Uridylation by TUT4 and TUT7 Marks mRNA for Degradation. Cell *159*, 1365–1376.

Lim, J., Kim, D., Lee, Y., Ha, M., Lee, M., Yeo, J., Chang, H., Song, J., Ahn, K., and Kim, V.N. (2018). Mixed tailing by TENT4A and TENT4B shields mRNA from rapid deadenylation. Science *361*, 701–704.

Loedige, I., Stotz, M., Qamar, S., Kramer, K., Hennig, J., Schubert, T., Löffler, P., Längst, G., Merkl, R., Urlaub, H., et al. (2014). The NHL domain of BRAT is an RNA-binding domain that directly contacts the hunchback mRNA for regulation. Genes Dev. *28*, 749–764.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat Rev Mol Cell Biol *8*, 479–490.

Lykke-Andersen, S., and Jensen, T.H. (2015). Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. Nature Reviews Molecular Cell Biology *16*, 665–677.

Ma, X.M., and Blenis, J. (2009). Molecular mechanisms of mTOR-mediated translational control. Nat Rev Mol Cell Biol *10*, 307–318.

MacArthur, H., Bubunenko, M., Houston, D.W., and King, M.L. (1999). Xcat RNA is a translationally sequestered germ plasm component in Xenopus. Mechanisms of Development *84*, 75–88.

Maillet, L., and Collart, M.A. (2002). Interaction between Not1p, a Component of the Ccr4-Not Complex, a Global Regulator of Transcription, and Dhh1p, a Putative RNA Helicase. J. Biol. Chem. *277*, 2835–2842.

Mayya, V.K., and Duchaine, T.F. (2019). Ciphers and Executioners: How 3′-Untranslated Regions Determine the Fate of Messenger RNAs. Front. Genet. *10*.

McLaren, A. (2003). Primordial germ cells in the mouse. Developmental Biology *262*, 1–15.

McMahon, A.C., Rahman, R., Jin, H., Shen, J.L., Fieldsend, A., Luo, W., and Rosbash, M. (2016). TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. Cell *165*, 742–753.

Mehlich, D., Garbicz, F., and Włodarski, P.K. (2018). The emerging roles of the polycistronic miR-106b~25 cluster in cancer – A comprehensive review. Biomedicine & Pharmacotherapy *107*, 1183–1195.

Meisenheimer, K.M., and Koch, T.H. (1997). Photocross-Linking of Nucleic Acids to Associated Proteins. Critical Reviews in Biochemistry and Molecular Biology *32*, 101–140.

Meng, X., Lindahl, M., Hyvönen, M.E., Parvinen, M., Rooij, D.G. de, Hess, M.W., Raatikainen-Ahokas, A., Sainio, K., Rauvala, H., Lakso, M., et al. (2000). Regulation of Cell Fate Decision of Undifferentiated Spermatogonia by GDNF. Science *287*, 1489–1493.

Meyer, S., Temme, C., and Wahle, E. (2004). Messenger RNA Turnover in Eukaryotes: Pathways and Enzymes. Critical Reviews in Biochemistry and Molecular Biology *39*, 197–216.

Mili, S., and Steitz, J.A. (2004). Evidence for reassociation of RNA-binding proteins after cell lysis: Implications for the interpretation of immunoprecipitation analyses. RNA *10*, 1692–1694.

Moore, M.J. (2005). From Birth to Death: The Complex Lives of Eukaryotic mRNAs. Science *309*, 1514–1518.

Morgan, M., Much, C., DiGiacomo, M., Azzi, C., Ivanova, I., Vitsios, D.M., Pistolic, J., Collier, P., Moreira, P.N., Benes, V., et al. (2017). mRNA 3′ uridylation and poly(A) tail length sculpt the mammalian maternal transcriptome. Nature *548*, 347–351.

Morimoto, H., Iwata, K., Ogonuki, N., Inoue, K., Atsuo, O., Kanatsu-Shinohara, M., Morimoto, T., Yabe-Nishimura, C., and Shinohara, T. (2013). ROS Are Required for Mouse Spermatogonial Stem Cell Self-Renewal. Cell Stem Cell *12*, 774–786.

Morimoto, H., Kanatsu-Shinohara, M., and Shinohara, T. (2015). ROS-Generating Oxidase Nox3 Regulates the Self-Renewal of Mouse Spermatogonial Stem Cells1. Biology of Reproduction *92*.

Morita, M., Suzuki, T., Nakamura, T., Yokoyama, K., Miyasaka, T., and Yamamoto, T. (2007). Depletion of Mammalian CCR4b Deadenylase Triggers Elevation of the p27Kip1 mRNA Level and Impairs Cell Growth. Molecular and Cellular Biology *27*, 4980–4990.

Much, C., Auchynnikava, T., Pavlinic, D., Buness, A., Rappsilber, J., Benes, V., Allshire, R., and O'Carroll, D. (2016). Endogenous Mouse Dicer Is an Exclusively Cytoplasmic Protein. PLOS Genetics *12*, e1006095.

Muraro, N.I., Weston, A.J., Gerber, A.P., Luschnig, S., Moffat, K.G., and Baines, R.A. (2008). Pumilio Binds para mRNA and Requires Nanos and Brat to Regulate Sodium Current in Drosophila Motoneurons. Journal of Neuroscience *28*, 2099–2109.

Nakagawa, T., Nabeshima, Y., and Yoshida, S. (2007). Functional Identification of the Actual and Potential Stem Cell Compartments in Mouse Spermatogenesis. Developmental Cell *12*, 195–206.

Nakagawa, T., Sharma, M., Nabeshima, Y., Braun, R.E., and Yoshida, S. (2010). Functional Hierarchy and Reversibility Within the Murine Spermatogenic Stem Cell Compartment. Science *328*, 62–67.

Nakamura, S., Kobayashi, K., Nishimura, T., Higashijima, S., and Tanaka, M. (2010). Identification of Germline Stem Cells in the Ovary of the Teleost Medaka. Science *328*, 1561–1563.

Nakamura, T., Yao, R., Ogawa, T., Suzuki, T., Ito, C., Tsunekawa, N., Inoue, K., Ajima, R., Miyasaka, T., Yoshida, Y., et al. (2004). Oligo-asthenoteratozoospermia in mice lacking Cnot7, a regulator of retinoid X receptor beta. Nature Genetics *36*, 528–533.

Naughton, C.K., Jain, S., Strickland, A.M., Gupta, A., and Milbrandt, J. (2006). Glial Cell-Line Derived Neurotrophic Factor-Mediated RET Signaling Regulates Spermatogonial Stem Cell Fate. Biol Reprod *74*, 314–321.

Neely, G.G., Kuba, K., Cammarato, A., Isobe, K., Amann, S., Zhang, L., Murata, M., Elmén, L., Gupta, V., Arora, S., et al. (2010). A Global In Vivo Drosophila RNAi Screen Identifies NOT3 as a Conserved Regulator of Heart Function. Cell *141*, 142–153.

Neumüller, R.A., Betschinger, J., Fischer, A., Bushati, N., Poernbacher, I., Mechtler, K., Cohen, S.M., and Knoblich, J.A. (2008). Mei-P26 regulates microRNAs and cell growth in the *Drosophila* ovarian stem cell lineage. Nature *454*, 241–245.

Ni, F.-D., Hao, S.-L., and Yang, W.-X. (2019). Multiple signaling pathways in Sertoli cells: recent findings in spermatogenesis. Cell Death Dis *10*, 1–15.

Niranjanakumari, S., Lasda, E., Brazas, R., and Garcia-Blanco, M.A. (2002). Reversible cross-linking combined with immunoprecipitation to study RNA–protein interactions in vivo. Methods *26*, 182–190.

Norbury, C.J. (2013). Cytoplasmic RNA: a case of the tail wagging the dog. Nature Reviews Molecular Cell Biology *14*, 643–653.

Oakberg, E.F. (1971). Spermatogonial stem-cell renewal in the mouse.

Oatley, J.M., Avarbock, M.R., Telaranta, A.I., Fearon, D.T., and Brinster, R.L. (2006). Identifying genes important for spermatogonial stem cell self-renewal and survival. Proceedings of the National Academy of Sciences *103*, 9524–9529.

Oatley, J.M., Avarbock, M.R., and Brinster, R.L. (2007). Glial Cell Line-derived Neurotrophic Factor Regulation of Genes Essential for Self-renewal of Mouse Spermatogonial Stem Cells Is Dependent on Src Family Kinase Signaling. J. Biol. Chem. *282*, 25842–25851.

Oatley, M.J., Kaucher, A.V., Racicot, K.E., and Oatley, J.M. (2011). Inhibitor of DNA Binding 4 Is Expressed Selectively by Single Spermatogonia in the Male Germline and Regulates the Self-Renewal of Spermatogonial Stem Cells in Mice. Biol Reprod *85*, 347–356.

Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S.C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A., et al. (2005).

Blimp1 is a critical determinant of the germ cell lineage in mice. Nature *436*, 207–213.

Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., and Saitou, M. (2009). A Signaling Principle for the Specification of the Germ Cell Lineage in Mice. Cell *137*, 571–584.

Otsuka, H., Fukao, A., Funakami, Y., Duncan, K.E., and Fujiwara, T. (2019). Emerging Evidence of Translational Control by AU-Rich Element-Binding Proteins. Front. Genet. *10*.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14, 417– 419.

Pauklin, S., Pedersen, R.A., and Vallier, L. (2011). Mouse pluripotent stem cells at a glance. Journal of Cell Science *124*, 3727–3732.

Peng, S.S.-Y., Chen, C.-Y.A., Xu, N., and Shyu, A.-B. (1998). RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. The EMBO Journal *17*, 3461–3470.

Phillips, B.T., Gassei, K., and Orwig, K.E. (2010). Spermatogonial stem cell regulation and spermatogenesis. Phil. Trans. R. Soc. B *365*, 1663–1678.

Plass, M., Rasmussen, S.H., and Krogh, A. (2017). Highly accessible AU-rich regions in 3' untranslated regions are hotspots for binding of regulatory factors. PLOS Computational Biology *13*, e1005460.

Pointud, J.-C. (2003). The intracellular localisation of TAF7L, a paralogue of transcription factor TFIID subunit TAF7, is developmentally regulated during male germ-cell differentiation. Journal of Cell Science *116*, 1847–1858.

Rahman, R., Xu, W., Jin, H., and Rosbash, M. (2018). Identification of RNA-binding protein targets with HyperTRIBE. Nature Protocols *13*, 1829–1849.

Raisch, T., Bhandari, D., Sabath, K., Helms, S., Valkov, E., Weichenrieder, O., and Izaurralde, E. (2016). Distinct modes of recruitment of the CCR4-NOT complex by Drosophila and vertebrate Nanos. EMBO J. *35*, 974–990.

Ramanathan, M., Porter, D.F., and Khavari, P.A. (2019). Methods to study RNA–protein interactions. Nature Methods *16*, 225.

Rato, L., Alves, M.G., Socorro, S., Duarte, A.I., Cavaco, J.E., and Oliveira, P.F. (2012). Metabolic regulation is important for spermatogenesis. Nature Reviews Urology *9*, 330–338.

Reid, D.W., Shenolikar, S., and Nicchitta, C.V. (2015). Simple and inexpensive ribosome profiling analysis of mRNA translation. Methods *91*, 69–74.

Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi, L., Meyer, M., Wong, J., Xu, C., et al. (2019). Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat Protoc *14*, 482–517.

Rissland, O.S., and Norbury, C.J. (2009). 3′ uridylation precedes decapping in a novel pathway of bulk mRNA turnover. Nat Struct Mol Biol *16*, 616–623.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24–26.

Rooij, D.G. de (2017). The nature and dynamics of spermatogonial stem cells. Development *144*, 3022–3030.

Rooij, D.G.D., and Russell, L.D. (2000). All You Wanted to Know About Spermatogonia but Were Afraid to Ask. Journal of Andrology *21*, 776–798.

Russell, L. D., Ettlin, R. A., Hikim, A. P. S. and Clegg, E. D. (1990). Histological and Histopathological Evaluation of the Testis. Clearwater, Fl: Cache River Press.

Saba, R., Kato, Y., and Saga, Y. (2014). NANOS2 promotes male germ cell development independent of meiosis suppression. Developmental Biology *385*, 32–40.

Sada, A., Suzuki, A., Suzuki, H., and Saga, Y. (2009). The RNA-Binding Protein NANOS2 Is Required to Maintain Murine Spermatogonial Stem Cells. Science *325*, 1394–1398.

Saga, Y. (2008). Mouse germ cell development during embryogenesis. Current Opinion in Genetics & Development *18*, 337–341.

Santos, M.G., Machado, A.Z., Martins, C.N., Domenice, S., Costa, E.M.F., Nishi, M.Y., Ferraz-de-Souza, B., Jorge, S.A.C., Pereira, C.A., Soardi, F.C., et al. (2014). Homozygous Inactivating Mutation in *NANOS3* in Two Sisters with Primary Ovarian Insufficiency. BioMed Research International *2014*, 1–8.

Sasaki, K., Yokobayashi, S., Nakamura, T., Okamoto, I., Yabuta, Y., Kurimoto, K., Ohta, H., Moritoki, Y., Iwatani, C., Tsuchiya, H., et al. (2015). Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. Cell Stem Cell *17*, 178–194.

Sato, K., Hayashi, Y., Ninomiya, Y., Shigenobu, S., Arita, K., Mukai, M., and Kobayashi, S. (2007). Maternal Nanos represses hid/skl-dependent apoptosis to maintain the germ line in Drosophila embryos. Proceedings of the National Academy of Sciences *104*, 7455–7460.

Saunders, L.R., and Barber, G.N. (2003). The dsRNA binding protein family: critical roles, diverse cellular functions. The FASEB Journal *17*, 961–983.

Saxton, R.A., and Sabatini, D.M. (2017). mTOR Signaling in Growth, Metabolism, and Disease. Cell *168*, 960–976.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Sha, Q.-Q., Yu, J.-L., Guo, J.-X., Dai, X.-X., Jiang, J.-C., Zhang, Y.-L., Yu, C., Ji, S.-Y., Jiang, Y., Zhang, S.-Y., et al. (2018). CNOT6L couples the selective degradation of maternal transcripts to meiotic cell cycle progression in mouse oocyte. The EMBO Journal *37*, e99333.

Shchepachev, V., Bresson, S., Spanos, C., Petfalski, E., Fischer, L., Rappsilber, J., and Tollervey, D. (2019). Defining the RNA interactome by total RNA-associated protein purification. Mol Syst Biol *15*.

Shirai, Y.-T., Suzuki, T., Morita, M., Takahashi, A., and Yamamoto, T. (2014). Multifunctional roles of the mammalian CCR4–NOT complex in physiological phenomena. Front Genet *5*.

Singh, G., Pratt, G., Yeo, G.W., and Moore, M.J. (2015). The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. Annu. Rev. Biochem. *84*, 325–354.

Smedley, D. *et al.* (2015). The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res* **43**, W589–W598.

Sonoda, J., and Wharton, R.P. (1999). Recruitment of Nanos to hunchback mRNA by Pumilio. Genes Dev. *13*, 2704–2712.

Steensel, B. van, and Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. Nature Biotechnology *18*, 424.

Stoecklin, G., Colombi, M., Raineri, I., Leuenberger, S., Mallaun, M., Schmidlin, M., Gross, B., Lu, M., Kitamura, T., and Moroni, C. (2002).

Functional cloning of BRF1, a regulator of ARE-dependent mRNA turnover. The EMBO Journal *21*, 4709–4718.

Subramaniam, K., and Seydoux, G. (1999). nos-1 and nos-2, two genes related to Drosophila nanos, regulate primordial germ cell development and survival in Caenorhabditis elegans. Development *126*, 4861–4871.

Sugimoto, Y., König, J., Hussain, S., Zupan, B., Curk, T., Frye, M., and Ule, J. (2012). Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. Genome Biology *13*, R67.

Sun, F., Xu, Q., Zhao, D., and Degui Chen, C. (2015). Id4 Marks Spermatogonial Stem Cells in the Mouse Testis. Scientific Reports *5*, 17594.

Suzuki, A., and Saga, Y. (2008). Nanos2 suppresses meiosis and promotes male germ cell differentiation. Genes Dev. *22*, 430–435.

Suzuki, A., Tsuda, M., and Saga, Y. (2007). Functional redundancy among Nanos proteins and a distinct role of Nanos2 during male germ cell development. Development *134*, 77–83.

Suzuki, A., Igarashi, K., Aisaki, K., Kanno, J., and Saga, Y. (2010). NANOS2 interacts with the CCR4-NOT deadenylation complex and leads to suppression of specific RNAs. PNAS *107*, 3594–3599.

Suzuki, A., Saba, R., Miyoshi, K., Morita, Y., and Saga, Y. (2012a). Interaction between NANOS2 and the CCR4-NOT Deadenylation Complex Is Essential for Male Germ Cell Development in Mouse. PLOS ONE *7*, e33558.

Suzuki, A., Niimi, Y., Shinmyozu, K., Zhou, Z., Kiso, M., and Saga, Y. (2016). Dead end1 is an essential partner of NANOS2 for selective binding of target RNAs in male germ cell development. EMBO Reports *17*, 37–46.

Suzuki, H., Sada, A., Yoshida, S., and Saga, Y. (2009). The heterogeneity of spermatogonia is revealed by their topology and expression of marker proteins

including the germ cell-specific proteins Nanos2 and Nanos3. Developmental Biology *336*, 222–231.

Suzuki, H., Ahn, H.W., Chu, T., Bowden, W., Gassei, K., Orwig, K., and Rajkovic, A. (2012b). SOHLH1 and SOHLH2 coordinate spermatogonial differentiation. Dev Biol *361*, 301–312.

Swartz, S.Z., Reich, A.M., Oulhen, N., Raz, T., Milos, P.M., Campanale, J.P., Hamdoun, A., and Wessel, G.M. (2014). Deadenylase depletion protects inherited mRNAs in primordial germ cells. Development *141*, 3134–3142.

Tagelenbosch, R.A.J., and de Rooij, D.G. (1993). A quantitative study of spermatogonial multiplication and stem cell renewal in the C3H/101 F1 hybrid mouse. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis *290*, 193–200.

Takashima, S., and Shinohara, T. (2018). Culture and transplantation of spermatogonial stem cells. Stem Cell Research *29*, 46–55.

Tebaldi, T., Zuccotti, P., Peroni, D., Köhn, M., Gasperini, L., Potrich, V., Bonazza, V., Dudnakova, T., Rossi, A., Sanguinetti, G., et al. (2018). HuD Is a Neural Translation Enhancer Acting on mTORC1-Responsive Genes and Counteracted by the Y3 Small Non-coding RNA. Molecular Cell *71*, 256-270.e10.

Tenenbaum, S.A., Carson, C.C., Lager, P.J., and Keene, J.D. (2000). Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. PNAS *97*, 14085–14090.

The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. Nucleic Acids Res *45*, D158–D169.

Tian, L., Liu, J., Xia, G.-H., and Chen, B.-A. (2017). RNAi-mediated knockdown of MCM7 gene on CML cells and its therapeutic potential for leukemia. Med Oncol *34*, 21.

Tsialikas, J., and Romer-Seibert, J. (2015). LIN28: roles and regulation in development and beyond. Development *142*, 2397–2404.

Tsuda, M., Sasaoka, Y., Kiso, M., Abe, K., Haraguchi, S., Kobayashi, S., and Saga, Y. (2003). Conserved Role of nanos Proteins in Germ Cell Development. Science *301*, 1239–1241.

Tyagi, G., Carnes, K., Morrow, C., Kostereva, N.V., Ekman, G.C., Meling, D.D., Hostetler, C., Griswold, M., Murphy, K.M., Hess, R.A., et al. (2009). Loss of Etv5 decreases proliferation and RET levels in neonatal mouse testicular germ cells and causes an abnormal first wave of spermatogenesis. Biol Reprod *81*, 258–266.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods *13*, 731–740.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP Identifies Nova-Regulated RNA Networks in the Brain. Science *302*, 1212–1215.

Ule, J., Jensen, K., Mele, A., and Darnell, R.B. (2005). CLIP: A method for identifying protein–RNA interaction sites in living cells. Methods *37*, 376–386.

Urdaneta, E.C., Vieira-Vieira, C.H., Hick, T., Wessels, H.-H., Figini, D., Moschall, R., Medenbach, J., Ohler, U., Granneman, S., Selbach, M., et al. (2019). Purification of cross-linked RNA-protein complexes by phenol-toluol extraction. Nat Commun *10*, 990.

Urlaub, H., Hartmuth, K., Kostka, S., Grelle, G., and Lührmann, R. (2000). A General Approach for Identification of RNA-Protein Cross-linking Sites within Native Human Spliceosomal Small Nuclear Ribonucleoproteins (snRNPs): ANALYSIS OF RNA-PROTEIN CONTACTS IN NATIVE U1 AND U4/U6.U5 snRNPs. J. Biol. Chem. *275*, 41458–41468.

Vakilian, H., Mirzaei, M., Sharifi Tabar, M., Pooyan, P., Habibi Rezaee, L., Parker, L., Haynes, P.A., Gourabi, H., Baharvand, H., and Salekdeh, G.H. (2015). DDX3Y, a Male-Specific Region of Y Chromosome Gene, May Modulate Neuronal Differentiation. J. Proteome Res. *14*, 3474–3483.

Wahle, E., and Winkler, G.S. (2013). RNA decay machines: Deadenylation by the Ccr4–Not and Pan2–Pan3 complexes. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1829*, 561–570.

Wang, Z., and Lin, H. (2004). Nanos Maintains Germline Stem Cell Self-Renewal by Preventing Differentiation. Science *303*, 2016–2019.

Wang, C., Wang, Z., Xiong, Z., Dai, H., Zou, Z., Jia, C., Bai, X., and Chen, Z. (2016). mTORC1 Activation Promotes Spermatogonial Differentiation and Causes Subfertility in Mice. Biology of Reproduction *95*, 97–97.

Wang, H., Morita, M., Yang, X., Suzuki, T., Yang, W., Wang, J., Ito, K., Wang, Q., Zhao, C., Bartlam, M., et al. (2010). Crystal structure of the human CNOT6L nuclease domain reveals strict poly(A) substrate specificity. The EMBO Journal *29*, 2566–2576.

Webb, S., Hector, R.D., Kudla, G., and Granneman, S. (2014). PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. Genome Biol *15*, R8.

Weber, S., Eckert, D., Nettersheim, D., Gillis, A.J.M., Schäfer, S., Kuckenberg, P., Ehlermann, J., Werling, U., Biermann, K., Looijenga, L.H.J., et al. (2010).

Critical Function of AP-2gamma/TCFAP2C in Mouse Embryonic Germ Cell Maintenance. Biol Reprod *82*, 214–223.

Webster, M.W., Chen, Y.-H., Stowell, J.A.W., Alhusaini, N., Sweet, T., Graveley, B.R., Coller, J., and Passmore, L.A. (2018). mRNA Deadenylation Is Coupled to Translation Rates by the Differential Activities of Ccr4-Not Nucleases. Molecular Cell *70*, 1089-1100.e8.

Weidmann, C.A., Qiu, C., Arvola, R.M., Lou, T.-F., Killingsworth, J., Campbell, Z.T., Hall, T.M.T., and Goldstrohm, A.C. (2016). Drosophila Nanos acts as a molecular clamp that modulates the RNA-binding and repression activities of Pumilio. ELife Sciences *5*, e17096.

Wharton, R.P., and Struhl, G. (1991). RNA regulatory elements mediate control of Drosophila body pattern by the posterior morphogen nanos. Cell *67*, 955–967.

Winkler, G.S., and Balacco, D.L. (2013). Heterogeneity and complexity within the nuclease module of the Ccr4-Not complex. Front. Genet. *4*.

Wittenberger, M.D., Hagerman, R.J., Sherman, S.L., McConkie-Rosell, A., Welt, C.K., Rebar, R.W., Corrigan, E.C., Simpson, J.L., and Nelson, L.M. (2007). The FMR1 premutation and reproduction. Fertility and Sterility *87*, 456–465.

Worthington, M.T., Pelo, J.W., Sachedina, M.A., Applegate, J.L., Arseneau, K.O., and Pizarro, T.T. (2002). RNA Binding Properties of the AU-rich Element-binding Recombinant Nup475/TIS11/Tristetraprolin Protein. J. Biol. Chem. *277*, 48558–48564.

Wreden, C., Verrotti, A.C., Schisa, J.A., Lieberfarb, M.E., and Strickland, S. Nanos and pumilio establish embryonic polarity in Drosophila by promoting posterior deadenylation of hunchback mRNA. 9.

Wu, X., Wang, B., Dong, Z., Zhou, S., Liu, Z., Shi, G., Cao, Y., and Xu, Y. (2013). A*NANOS3*mutation linked to protein degradation causes premature ovarian insufficiency. Cell Death & Disease *4*, e825.

Xu, W., Rahman, R., and Rosbash, M. (2018). Mechanistic implications of enhanced editing by a HyperTRIBE RNA-binding protein. RNA *24*, 173–182.

Yamaji, M., Seki, Y., Kurimoto, K., Yabuta, Y., Yuasa, M., Shigeta, M., Yamanaka, K., Ohinata, Y., and Saitou, M. (2008). Critical function of *Prdm14* for the establishment of the germ cell lineage in mice. Nature Genetics *40*, 1016–1022.

Yamashita, A., Chang, T.-C., Yamashita, Y., Zhu, W., Zhong, Z., Chen, C.-Y.A., and Shyu, A.-B. (2005). Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. Nat Struct Mol Biol *12*, 1054–1063.

Ye, B., Petritsch, C., Clark, I.E., Gavis, E.R., Jan, L.Y., and Jan, Y.N. (2004). nanos and pumilio Are Essential for Dendrite Morphogenesis in Drosophila Peripheral Neurons. Current Biology *14*, 314–321.

Yi, H., Park, J., Ha, M., Lim, J., Chang, H., and Kim, V.N. (2018). PABP Cooperates with the CCR4-NOT Complex to Promote mRNA Deadenylation and Block Precocious Decay. Molecular Cell *70*, 1081-1088.e5.

Yoshida, S., Takakura, A., Ohbo, K., Abe, K., Wakabayashi, J., Yamamoto, M., Suda, T., and Nabeshima, Y. (2004). Neurogenin3 delineates the earliest stages of spermatogenesis in the mouse testis. Developmental Biology *269*, 447–458.

Yoshida, S., Sukeno, M., Nakagawa, T., Ohbo, K., Nagamatsu, G., Suda, T., and Nabeshima, Y. (2006). The first round of mouse spermatogenesis is a distinctive program that lacks the self-renewing spermatogonia stage. Development *133*, 1495–1505.

Yoshida, S., Nabeshima, Y.-I., and Nakagawa, T. (2007). Stem Cell Heterogeneity. Annals of the New York Academy of Sciences *1120*, 47–58.

Zhang, C., and Darnell, R.B. (2011). Mapping *in vivo* protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nature Biotechnology *29*, 607–614.

Zhang, T., and Zarkower, D. (2017). DMRT proteins and coordination of mammalian spermatogenesis. Stem Cell Research *24*, 195–202.

Zhang, F., Zhang, Y., Lv, X., Xu, B., Zhang, H., Yan, J., Li, H., and Wu, L. (2019). Evolution of an X-Linked miRNA Family Predominantly Expressed in Mammalian Male Germ Cells. Mol Biol Evol *36*, 663–678.

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. Molecular Cell *40*, 939–953.

Zhou, Z., Shirakawa, T., Ohbo, K., Sada, A., Wu, Q., Hasegawa, K., Saba, R., and Saga, Y. (2015). RNA Binding Protein Nanos2 Organizes Post-transcriptional Buffering System to Retain Primitive State of Mouse Spermatogonial Stem Cells. Developmental Cell *34*, 96–107.

Zhou, Z., Kawabe, H., Suzuki, A., Shinmyozu, K., and Saga, Y. (2017). NEDD4 controls spermatogonial stem cell homeostasis and stress response by regulating messenger ribonucleoprotein complexes. Nature Communications *8*, 15662.

Zwartjes, C.G.M., Jayne, S., Berg, D.L.C. van den, and Timmers, H.T.M. (2004). Repression of Promoter Activity by CNOT2, a Subunit of the Transcription Regulatory Ccr4-Not Complex. J. Biol. Chem. *279*, 10848–10854.